UNIVERSITAT ROVIRA I VIRGILI

PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES INDÍGENAS OAXAQUEÑAS

María Martínez Iglesias

UNIVERSITAT ROVIRA i VIRGILI

Enhancing the ELECTRE decision support method with semantic data

MIRIAM MARTÍNEZ GARCÍA

DOCTORAL THESIS
2018

UNIVERSITAT ROVIRA I VIRGILI

PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS

María Martínez Iglesias

Miriam Martínez García

Enhancing the ELECTRE decision support method with semantic data

Doctoral Thesis

Supervised by:
Dr. Aïda Valls
Dr. Antonio Moreno

Department of Computer Engineering and Mathematics

UNIVERSITAT ROVIRA i VIRGILI

Tarragona

2018

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

UNIVERSITAT ROVIRA i VIRGILI
Departament d'Enginyeria
Informàtica i Matemàtiques
Av. Països Catalans 26,
43007 Tarragona, Spain

We STATE that the present study, entitled "Enhancing the ELECTRE decision support method with semantic data", presented by Miriam Martínez García for the award of the degree of Doctor, has been carried out under our supervision at the Department of Computer Engineering and Mathematics of this university.

Tarragona, July, 28th 2018

Doctoral Thesis Supervisors

Dr. Aïda Valls Mateu

Dr. Antonio Moreno Ribas

UNIVERSITAT ROVIRA I VIRGILI

PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES INDÍGENAS OAXAQUEÑAS

María Martínez Iglesias

Acknowledgements

First of all, I am deeply grateful with God because he provides me the confidence and strength that I need in every moment of my life. Without a doubt, my family is and always will be in my heart and mind. I specially want to say thanks to a woman who support me in every step, my mom Victoria. Besides, my brother Fausto and sister Erandy to whom I have endless love and I have always been able to count on them. To my Father, thank you for all the support. During these years I have met lovely people for whom I keep great affection, my friends. Particularly a person who changed my life in a way that I never imagined, my husband Gildas, who with unending love and patience is by my side and encourage me to never give up. At last but not least, I want greatly to thank my mentors Aïda Valls and Antonio Moreno, because this project could not be possible without their continuous guidance and professionalism. Special thanks to Arantza Aldea from the Oxford Brookes University and René Bañares from Oxford University for the wide collaboration and kindness during my exchange program.

Also I want to say thank you a members of ITAKA group in special Luis Del Vasto, Joan Borràs, Sílvia Sanromà, Héctor Mercado and Jordi Pascual.

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

Abstract

Multi-Criteria Decision Aiding (MCDA) studies the way in which people reach a decision when options are defined on a set of criteria, with the aim of developing tools that help the decision maker. In our work, we consider outranking methods. This thesis is focused on the methodology ELECTRE (ELimination and Choice Expressing REality) that was, in fact, the first outranking method in MCDA. Outranking methods consider heterogeneous criteria to evaluate the performance of the alternatives and compare them, including different numerical and ordinal scales in the set of criteria. Nowadays it is becoming more common to find decision problems involving non-numerical information, such as multi-valued semantic criteria, which may take as values the concepts of a given domain ontology.

In this PhD Thesis, I propose a new way of handling semantic criteria to avoid the aggregation of the numerical scores before the ranking procedure. This method, called ELECTRE-SEM, follows the same principles than the classic ELECTRE but in this case the concordance and discordance indices are defined in terms of the pairwise comparison of the interest scores.

I also propose to create a semantic user profile by storing preference scores into the ontology. This preferential information may be later exploited to rank and recommend the most suitable alternatives for each user. The numerical interest score attached to the most specific concepts permits to distinguish better the preferences of the user, improving the quality of the decision by the incorporation of an aggregation method to infer the user's preferences from the taxonomic relations between concepts.

The proposed methodology has been applied in two case studies: the assessment of power generation plants and the recommendation of touristic activities in Tarragona.

Keywords: MCDA, ELECTRE, Preference Learning, Ontology, Semantic Data.

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

Contents

List of Figures

List of Tables

1

Introduction

In the current Knowledge Society in which we are living, there is an almost instantaneous access to huge amounts of data about any topic. This fact is a double-edged sword as, on the one hand, we can easily obtain information about any issue but, on the other hand, we find it increasingly difficult to cope with this deluge of data. In our daily life we are constantly confronted with many situations in which we must make a decision, choosing a certain option from a given set of alternatives. Given the amount of available information, it is common that there are hundreds or thousands of potential alternatives, and we may have dozens of data on each of them. Thus, it may be cognitively hard to analyze all these data, evaluate all the options and make the most appropriate decision [45]. One of the current areas of study in Artificial Intelligence (AI) is the one devoted to *Recommender Systems (RS)*, which are capable of analyzing automatically all the alternatives in order to show to the decision maker the most adequate ones.

In many situations each of the alternatives may be represented as a set of (attribute, value/s) pairs. *Multi-Criteria Decision Aiding (MCDA)* is an academic field devoted to the development of decision support methods in this setting. MCDA methods analyze all the alternatives and try to discover which of them suit better the needs or the preferences of the decision maker. Thus, in this field it is of paramount importance to have accurate and complete knowledge about these preferences.

Decisions based on multiple criteria are often very hard to make, since

usually there is not any alternative that is better than all the others in all the attributes (or criteria). Actually, it is very common that criteria are contradictory (e.g. criteria measuring the quality of an option vs criteria showing its economic cost). MCDA methods also have to take into account that criteria may take different kinds of values. The classic MCDA techniques usually deal with numerical and categorical criteria; however, most of the information that we can access through the Internet is in textual form (e.g. the tourist attractions in a given destination). Thus, a current challenge in the field of MCDA is how to incorporate the analysis of this kind of information in the current methodologies. In particular, it is necessary to study how to represent the preferences of the user on this type of data and how to use this preferential information when evaluating the alternatives. This is the main challenge considered in this dissertation.

There are different kinds of MCDA methodologies, which may be based on decision rules [28], utility functions [45] or relational models. In this work we have focused on this latter category, also known as *outranking* methods; in particular, we have centered our attention on the ELECTRE family of methods [73]. Outranking methods have been very succesful and they have been widely applied in many different domains.These techniques construct a preference structure from the pairwise comparison of the alternatives, which is based on two ideas inspired by voting theory: concordance (or "the choice of the majority") and discordance (or "the respect to minorities"). An important advantage of these methods is that they can work directly with purely ordinal scales, without requiring their transformation into abstract ones with an arbitrary range. A second advantage is that indifference and preference thresholds can be used to model the uncertainty of the decision maker when analyzing the alternatives. However, one of the main shortcomings of ELECTRE is that alternatives can only be defined in terms of numerical and ordinal criteria, so the incorporation of other types of data is a current area of research.

In a nutshell, this dissertation studies how to incorporate textual information in the classic ELECTRE MCDA methodology. Thus, we have to study how to represent this kind of information in the alternatives, how to represent the preferences of the decision maker on the values of this kind of criteria in the user profile, and how to incorporate the analysis of these data into ELECTRE. It is important to note that we do not want

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

3

to represent and analyze this information in a purely textual way. It is necessary to understand the *meaning* of the values to assess if an alternative is close to the preferences of the user or to determine if two alternatives are similar. For example, if we are analyzing the tourist attractions in a city, we have to understand that "Amphitheatre" is a kind of "Roman Building", so it is appropriate to recommend the city to a tourist interested in this kind of architecture. If a city has a "Museum of Modern Art" and another one has a "Modern Art Exhibition", we have to understand that both activities are related to Modern Art and the two cities may be similar in this respect.

In order to work at the conceptual level, rather than at the syntactic level, we have defined the notion of a *semantic* criterion. The value that an alternative may take on a criterion of this kind is a *concept*. A common way of representing conceptual information in Artificial Intelligence is to use ontologies. An *ontology* is basically formed by a set of concepts, which may be linked through *taxonomic* (i.e. *is-a*) and *non-taxonomic* relationships. It can also store properties of the concepts and specific instances of them. In this work there will be a domain ontology for each semantic criterion, which will specify the values that it can take and the taxonomic relationships between them. This taxonomical information will be used to evaluate the semantic similarity between two concepts. An extra complexity is added by the fact that we are going to consider that semantic criteria are usually *multi-valued*, so a given alternative will have a list of values in each semantic citerion, and not a single value.

In summary, the main research questions leading to the research shown in this dissertation are the following:

- How can we represent in an individual user profile the information about the preferences of the user with respect to the values that a semantic criterion may take?

- Taking into account that a common domain ontology may have hundreds or thousands of concepts, how can we obtained a detailed and complete account of the preferences of the user with respect to all of them, without forcing the user to fill an exhaustive questionnaire?

- Even if we had complete information about the preferences of the user with respect to the values that a semantic attribute may take,

how can we improve the current ELECTRE methods so that they can take into account this preferential information?

- How do we handle the additional complexity of dealing with multi-valued semantic attributes? In an outranking setting, how can we compare the lists of concepts associated to two alternatives in a given semantic criterion?

The answers to all those questions have led to the development of a new method for the ELECTRE family, which we have called ELECTRE-SEM. This new MCDA framework permits the analysis of alternatives which contain not only numerical or ordinal attributes but also multi-valued semantic ones. We have applied it to two real world domains (recommendation of touristic activities and analysis of power generation plants), as reported later in this document. In the following sections we detail the thesis objectives and the academic contributions of this work.

1.1   Objectives of the thesis

The objectives of this Ph.D. thesis can be summarized as follows:

1. The first goal of the thesis is to improve the management of preferences in ontology-based recommender systems. This objective can be divided in two sub-objectives: to define a way to represent the partial and incomplete information about the preferences of a user on the values of a semantic attribute in an ontology, and to design an algorithm that may complete such a partial set of preferences, taking into account the known ones and the structure of the underlying ontology.

2. The second goal of the dissertation is to improve the classical ELECTRE multi-criteria decision aid methodology so that it can be applied in problems in which the decision alternatives are not only defined in terms of numerical or categorical attributes but also multi-valued semantic ones.

3. The third goal of the work is to evaluate the new enhanced semantic ELECTRE methodology in real scenarios, to assess its usefulness and validity in different circumstances.

1.2   Thesis contributions

The contributions of this Ph.D. thesis can be summarized as follows:

1. In order to complete the first objective, two tasks have been carried out. In the first one, we have defined a way of representing the information about the preferences of a user on the different values that a semantic attribute may take, which are the leaves of a domain ontology. Concretely, there will be a numerical *Tag Interest Score* for each of these specific concepts in the user profile. As the initial information that a recommender system has on the user's preferences will probably be partial, in the second task we have designed and implemented a new algorithm that is able to infer the preferential information on all the leaves of the domain ontology. Given a concept c, this algorithm finds out which are the ontology concepts that are most semantically similar to c and have a known interest score, and then combines these scores to find out the preference on c, taking into account the structure of the ontology. The aggregation of the preferences is made using a Weighted Ordered Weighted Aggregation (WOWA), so it is a flexible approach in which many different aggregation policies may be considered. This preference learning algorithm has been compared with other similar methods in the literature, obtaining a good performance for different settings of its parameters.

    These contributions are described in the next chapter of this dissertation. They have been presented in the following papers:

    - Martínez-García, M., Valls, A., Moreno A. *Using aggregation operators to infer semantic preferences.* 28th European Conference on Operational Research (EURO conference), pp.248, Poznan, Poland, July 2016.
    - Martínez-García, M., Valls, A., Moreno, A. *Inferring preferences in ontology-based recommender systems using WOWA.* Journal of Intelligent Information Systems (submitted, second review).

2. Concerning the second objective, the ELECTRE method has been extended, giving rise to a the new system *ELECTRE-SEM*. More

concretely, we have defined new concordance and discordance indices for semantic criteria. They are based on the concept of *Semantic Win Rate*, which is a normalised measure of how good is an alternative with respect to another taking into account a certain multi-valued semantic criterion. By applying these concordance and discordance indices it is now possible to apply the ELECTRE methodology to situations in which criteria are not only numerical and categorical but also semantic. This new methodology of the ELECTRE family is described in chapter 4 of this dissertation. It is illustrated with a case study based on the recommendation of touristic activities.

ELECTRE-SEM was described in the following paper:

- M.Martínez-García, A.Valls, A.Moreno. *Construction of an outranking relation based on semantic criteria with ELECTRE-III.* Information Processing and Management of Uncertainty in Knowledge-Based Systems, Part II. Eds: J.P.Carvalho, M.J.Lesot, U.Kaymak, S.Vieira, B. Bouchon-Menier, R.Yager. Communications in Computer and Information Science (CCIS) 611, pp. 238-249. Springer- Verlag, 20168. ISBN 978-3-319-40580-3.

Preliminary versions of this work were also presented in the following specialised international meetings:

- Martínez-García, M., Valls, A., Moreno, A., *Making decisions with ELECTRE from semantic data.* European Working Group on Multicriteria Decision Aiding workshop (EWG-MCDA), Odense, Denmark, September 2015.
- Martínez-García, M., Valls, A., Moreno A., *New concordance and discordance indices in ELECTRE-III for semantic data.* European Working Group on Multicriteria Decision Aiding workshop (EWG-MCDA), Barcelona, Spain 2016.

3. Finally, the new ELECTRE-SEM multi-criteria decision aid framework has been applied in two real world domains. The first one is the recommendation of tourism and leisure activities in the area of Tarragona. This work has been made in close collaboration with the Scientific and Technological Park of Tourism and Leisure (Vila-Seca,

Tarragona). The second domain of interest is the evaluation of electricity generation technologies, taking into account economic and environmental criteria. This work was mostly developed in the stay of 3 months made at the research group of Dr. Arantza Aldea at the School of Engineering, Computing and Mathematics of Oxford Brookes University (Oxford, UK).

These two applications are described in chapter 5 of this dissertation, and they have also been reported in the following publications:

- Martínez-García, M., Valls, A., Moreno, A., Aldea, A., *A semantic multi-criteria approach to evaluate different types of energy generation technologies.* Environmental Modelling & Software, Elsevier 2018 (accepted, in press).
  DOI: 10.1016/j.envsoft.2018.04.003.

- Martínez-García, M., Valls, A., Moreno, A. *On the use of semantic criteria in ELECTRE in a touristic recommender system.* 29th European Conference on Operational Research (EURO-2018), Valencia (Spain), July 2018.

1.3   Organization of this manuscript

The rest of this dissertation is divided into the following chapters:

- Chapter 2 - Representing and inferring preferences in ontology-based recommender systems.

  This chapter starts with a review of the state of the art on ontology-based recommender systems, especially in the Tourism field. It explains how to make a (possibly partial) numerical representation of the preferences of the user on the values of a semantic attribute using *Tag Interest Scores*. It then describes how to complete all the preferential information of the user by leveraging semantic similarity measures (path length) and aggregation operators (Weighted Ordered Weighted Aggregation). Some experiments using Leisure and Sport ontologies are reported. Finally, the new preference learning approach is compared with three methods from the state of the art.

- Chapter 3 - Multiple Criteria Decision Aiding.

  This chapter provides a brief introduction to the field of Multi-Criteria Decision Aiding (MCDA). It is an academic area that studies ways of supporting a decision maker that has to choose/rank/sort a set of alternatives which are defined on different numerical/ordinal criteria. Three main categories of MCDA methods are presented: logic approaches (based on decision rules), functional methods (mainly based on multi-attribute utility theory) and outranking techniques (based on the pairwise comparison of alternatives, e.g. using the PROMETHEE framework). This dissertation has focused on the later category, more specifically in the ELECTRE family of methods.

- Chapter 4 - ELECTRE-SEM: an outranking method with semantic criteria.

  This chapter starts with a general presentation of the ELECTRE methodology, that aims to compare pairs of alternatives so that different kinds of relationships (indifference, preference and incomparability) can be defined among them. The comparison between two alternatives leads to the calculation of concordance and discordance indices, which are used to compute an overall outranking credibility value for each pair of options. These values are then used to obtain a partial preorder between the alternatives. In this chapter it is shown how to calculate the concordance and discordance indices for multi-valued semantic attributes, in order to incorporate this kind of information into the ELECTRE method. A case study analyzes how the different parameters of the methodology influence the way in which a set of touristic activities are ranked.

- Chapter 5- Applications.

  This chapter shows the application of the new ELECTRE-SEM methodology to two domains of interest: tourism and environment. First we apply the system to real data used in a recommender of touristic activities developed at the Scientific and Technological Park of Tourism and Leisure. After that, we show how the new framework may be used to analyze different alternatives for the generation of electricity, taking into account the actual values of some economic and environmental attributes.

- Chapter 6- Conclusions and future work.

  The final chapter summarizes the main contributions of the dissertation, reports the main conclusions drawn from this work and outlines some lines of future research.

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

2

Representing and inferring preferences in ontology-based recommender systems

Semantic (or ontology-based) recommender systems focus on the analysis of a set of alternatives defined on semantic attributes, in order to rank them or to select the ones that fit better with the user preferences [54]. A common option in this kind of recommenders is to use domain ontologies both to structure the possible values of each semantic attribute and to store, in some way, the preferences of the user on those values [4], [92]. Usually a semantic attribute may take as value a list of the most specific concepts of the ontology (i.e., the leaves of the tree), and the score of an alternative depends on the preference of the user with respect to those specific concepts.

This chapter starts with a brief review of the use of ontologies in recommender systems. After that we present a new way of representing preferential information on semantic attributes, using *Tag Interest Scores*. In real applications it may be assumed that the system would only have partial information about the preferences of the user; thus, we have designed and implemented a novel algorithm to estimate the missing preferences. At the end of this chapter we show an application of this algorithm to a touristic case study and we compare it with some methods of the state of the art.

2.1   Related Work

There are several ways in which ontologies have been used to represent the user profile. The simplest way is to associate to each user an explicit list of the concepts (attribute values) in which he/she is interested (e.g. [79], [7], [75], [49]). This option is not very informative, since it is not possible to express levels of interest on different concepts. A more interesting and widespread approach is to represent the preferences of a user with a vector of real-valued features, in which each position contains the degree of interest of the user with respect to a concept of the ontology (e.g. [87], [81], [16], [77], [42], [57], [86]). Some works also add a measure of the credibility associated to the information stored in the profile. The preference rating values may be uncertain because in many cases they are not fully provided in an explicit way by the user, but have to be inferred or discovered in some way. These confidence degrees associated to each concept may be later used as weighting factors in the recommendation process [23], [11].

Tourism is one of the fields in which ontology-based recommender systems have been most heavily applied in the last years [10], [6]. For example, in [30] a Tourism taxonomy was designed to categorize attractions in classes like 'Gothic Art', 'Museums' or 'Religious Buildings'. The Tourism ontology defined in [59] had properties like 'part of', 'hasQuality', 'location' or 'date'. The e-Tourism ontology defined in [30] also contained non-taxonomic properties like locatedIn, interestedIn or hasCurrency, that allowed the system to answer questions like which activities may be visited by a certain type of tourists, which is the location of interesting places or when they can be visited. Both [30] and [51] used explicit rules to be able to deduce information from the ontology and to answer queries on it. For instance, a rule like 'fact(? X type architecture 0.9*?N) :- fact(? X type church ?N)' states that if an item belongs to the church category with a score N, it can also be considered a member of the architecture category with a score 0.9*N. In the SigTur recommender system [60] the authors defined a 5-levels Tourism ontology with over 200 concepts. The first level of the ontology contained 8 general categories (Events, Nature, Culture, Leisure, Sports, Towns, Routes and Viewpoints), which were refined in the subsequent levels. In [94] the authors described a recommender of tourist destinations based on Bayesian networks, which used an ontology on user

2.1. Related Work 13

profiles and another one on touristic information. Some works consider fuzzy ontologies, in which a concept may be related to a certain degree with another one (e.g. [17] use a fuzzy ontology on wines to recommend the most appropriate wine in a particular context). More complex recommender systems consider a set of ontologies; for example, [49] presents a semantic-based Tourism information system that employs a network of ontologies, called ContOlogy, composed by 11 ontologies, 86 classes, 63 properties and 43 restrictions. These ontologies represent the information about visitors, preferences, roles, activities, environment, devices, network, motivations, location, time and Tourism objects.

One of the key problems in ontology-based recommender systems is the initialization of the user profile, i.e. the acquisition of the preferences of the user with respect to the possible values of each semantic criterion. This preferential information may be acquired explicitly at the beginning of the recommendation session, by asking the user to complete some kind of form, to answer a questionnaire or to rate some alternatives (e.g. [16], [79], [61], [8]). This approach provides precise information, since it is given directly by the user; however, it is an intrusive elicitation mechanism, and most users are not keen on spending time providing this information. Moreover, if the number of concepts in the ontologies associated to the semantic attributes is large, it is not feasible in practice to ask the user to express his/her preference on each concept. A possible solution consists on using some kind of spreading procedure to propagate the scores given by the user to some concepts to the rest of concepts of the ontology. For example, in the SigTur system [10] the user is initially asked to provide the preferences only on the top categories of the touristic ontology, and this information is spread to the subclasses (adding a certainty factor that decreases with the distance to these concepts). Another possibility is to try to learn the user preferences by analyzing his/her interaction with the system (e.g. the alternatives that are selected/viewed/deleted/purchased, the ratings given to the alternatives, or even the time spent with each alternative). SigTur [10] also employed this kind of techniques to update dynamically the information on the preferences of the user. The main advantage of this approach is that users do not need to spend time thinking about their preferences and making them explicit; however, more sophisticated computational approaches are required to try to understand the preferences

of the user, and this information may have some associated uncertainty.

This chapter is going to present a new way of completing a partial set of preferences in the initialization process of the user profile in an ontology-based recommender system, so a comprehensive study of related methods has been done. Among the works in the literature we can find different approaches to obtain the users preferences in ontology-based recommender systems. Table 2.1 shows their main distinguishing features. There are papers that do not explain the process followed to initialize the user profile. Others assume that the user will give explicitly the initial preference scores. Six methods require additional information, i.e. the recommenders whose purpose is to filter documents. In many cases the user profile is completed by analyzing the user's actions on the recommender system. There are only three works comparable to the setting considered in this paper (without additional sources of information and without any users feedback requirements), which are [16], [81] and [76]. In the experimental section they are compared with the method proposed in this chapter.

Regardless of the method employed to acquire the user preferences, it is very unlikely that the system can have reliable and complete information on the interest of the user on each possible value of each semantic attribute. Thus, in this work it is assumed that it is more realistic to expect that, given a certain semantic attribute, the system will initially only have information on the preferences of the user with respect to a small set of concepts of the ontology. This hypothesis motivates the need to have a computational mechanism that is able to complete the preferences of the user taking into account the initial partial information and the structure of the ontology.

In this chapter a method to store the user preferences by means of an ontology is proposed. First, we define the user profile structure (section 2.2). Assuming that the user will introduce manually an interest score only for a subset of the concepts of the ontology, an algorithm to estimate the interest of the user on the rest of the concepts is presented in section .

Table 2.1: Analysis of user profile initialization methods

Ref	Method of calculation	Additional sources	Implicit feedback
[5]	Given by user	None	Yes
[8]	Spreading by means of semantic associations parent-child	None	No
[10]	Downwards propagation of the interest scores given to the most general concepts	None	No
[16]	Constrained Spreading Activation (CSA)	None	No
[19]	Spreading activation	None	Yes, actions of users on the items
[24]	TF-IDF scheme	Documents and Open Directory Project	Yes, first documents searched
[62]	Train a learning model from examples of ratings	Linked Open Data	Yes
[87]	Given by the user	None	No
[36]	Spreading with contextual information	Context: time, place, people relationships, etc.	No
[42]	Bayesian probabilistic propagation model	History of records of the user's Web searches	Continuous recalculation with decay factor
[58]	Estimation from the K-nearest users	Age of the publications used and set of similar users	No
[57]	Not explained	None	Yes, with previously browsed papers
[61]	Transformation to AHP decision model	None	No
[76]	Average of scores of the closest concepts	None	No
[77]	Not explained	Web documents	Yes, web URL and its clicks
[79]	Given by the user	None	Yes, clicks to items
[81]	Spreading activation	None	Yes
[86]	All concepts with the lowest possible score	None	Yes, from user's ratings to items
[99]	Propagation only to the super-classes (ancestors) given scores of instances	None	No

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

16   **Chapter 2. Representing and inferring preferences in
ontology-based recommender systems**

2.2   Representing semantic preferences with Tag Interest Scores

In this work we use a feature-based approach, in which each user has a personal profile that consists on a numerical preference score associated to the set of all possible tags that may appear in the alternatives. This set of tags is restricted to the most specific terms of the ontology (the ones that do not have descendants). The ontology is structured using taxonomical relations (is-a) where multiple inheritance is possible. Therefore, the tags (or concepts) may have multiple parents. Fig. 2.1 shows a portion of an ontology that classifies different aquatic sports. Notice that Boating and Fishing are sport activities made in the river or in the sea and, hence, they are subclasses of these two concepts. Subtypes of boating and sailing are defined in this ontology, although they are not displayed in the figure.

The *Tag Interest Score TIS(t)* is a numerical score between 0 and 1 that indicates the satisfaction degree of the user with the corresponding tag t according to the decision maker's goals. The tag score may have two possible directions: maximization (1 is the best score) or minimization (0 is the worst score). The former is known as a gain criterion and the latter as a cost criterion. The direction of *TIS(t)* must be decided according to the decision problem to be solved. In some problems, the concepts of the ontology may indicate negative features, such as environmental pollutants, hence the *TIS(t)* should be minimized if it is associated to the quantity of the pollutant. On the contrary, in other problems it may indicate elements that the user is searching for. In this case, *TIS(t)* is usually related to the degree of interest on the concepts and it is positively treated. We assume that *TIS(t)* has to be maximized. In Fig. 2.1 we can see an example of a tourist's profile with some interest scores, like *TIS*(Kayaking) = 0.8, *TIS*(Rafting) = 0.7 or *TIS*(Windsurfing) = 0.3, and we can see that this tourist prefers sports activities in the river (with tags with the highest scores), and he/she does not like surfing sports, except maybe Banana Rafting. In this example, there are three leaves without score: Canoeing, ScubaDiving and Wakesurfing.
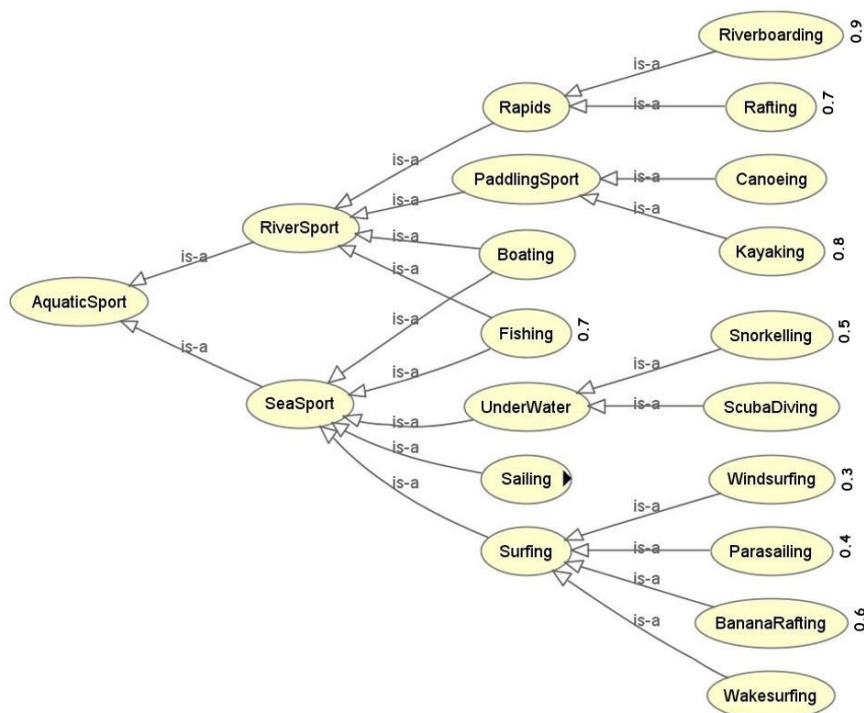
Figure 2.1: User profile ontology about aquatic sports with Tag Interest Scores

2.3   Preliminaries

Before presenting the proposed algorithm for the calculation of the missing preference scores in the ontology, we introduce in this section some tools that will be used. First, semantic similarity measures are presented because they will be used to find the related concepts in the ontology with known values (section 2.3.1). Second, the WOWA aggregation operator is defined in order to be later applied to merge the preference values of these neighbour concepts (section 2.3.2).

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

Chapter 2.   Representing and inferring preferences in
18   **ontology-based recommender systems**

2.3.1   Semantic similarity measures

Ontologies model the knowledge about the concepts in a certain domain using several types of relations, being the most common the taxonomical relations between a general concept and its sub-concepts (i.e. *is-a* relations). The exploitation of the information stored in ontologies is quite common in different fields, such as in Computational Linguistics for text analysis and text categorization, among others. In many of these tasks it is necessary to measure the semantic relatedness between two different concepts. Semantic similarity functions can be basically divided into two main categories: *distributional measures* and *ontology-based measures.* The distributional approaches use text corpora as the source to infer the semantics of the terms. They are based on the assumption that words with similar distributional properties have similar meanings [91]. Such measures take into account the co-occurrence of the words associated to the concepts in the same texts. The second approach relies on the relations between the concepts found in an ontology. Three types of ontology-based semantic similarity measures are distinguished [50]: edge-counting, feature vectors and information content. Edge-counting similarity functions use the number of edges separating two concepts to calculate the distance between them. The simplest measure is known as *Path Length* and it takes as similarity the minimum number of *is-a* links needed to connect two nodes of the ontology [68]. Such method to calculate the distance between terms has some weaknesses such as not considering the depth (i.e. the specificity) of the compared concepts. In this sense, other measures [96], [22] consider the depth of the concepts in the taxonomy, because concept specializations become less distinct the more they are refined. So, equally distant pairs of concepts belonging to an upper level of a taxonomy should be considered less similar than those belonging to a lower level. In case of multiple inheritance, it may be interesting to use a similarity measure that takes into account the number of common ancestors of the compared concepts, as proposed in [33]. Secondly, feature-based measures estimate the similarity according to other common semantic features between the two concepts, such as synonyms, meronyms or other semantic relationships [65]. Finally, a third approach consists on a conceptualization of information content of a term as the probability of its occurrence [69]. This probability can be computed from an external corpus or internally from the intrinsic information of the ontology structure [66].

2.3.2   The WOWA operator

The *Ordered Weighted Average* (OWA) is a flexible aggregator operator that admits different degrees of conjunction/disjunction [97]. This technique has been widely studied and used in many decision-making problems [98]. Before defining the OWA operator, which is the basis for the WOWA operator, some preliminary concepts are formalized.

Definition 1. *A vector $v = (v_1...v_n)$ is a weighting vector of dimension n if and only if $\forall_{i,n \geq i \geq 1}$, $v_i \in [0,1]$ and $\sum_{i=1}^{n} v_i = 1$.*

Definition 2. *A mapping $AM: \mathbb{R}^n \to \mathbb{R}$ is an arithmetic mean of dimension n if $AM(a_1, ..., a_n) = \left(\frac{1}{n}\right) \sum_{i=1}^{n} a_i$.*

Definition 3. *Let p be a weighting vector of dimension n; then, a mapping $WM_p: \mathbb{R}^n \to \mathbb{R}$ is a weighted mean of dimension n if $WM_p(a_1, ..., a_n) = \sum_{i=1}^{n} p_i a_i$.*

The OWA operator is defined as a linear combination of the data with respect to a weighting vector, similarly to the weighted mean. However, in this case, a permutation of the values that are aggregated $a_{\sigma(i)}$ plays a central role in the definition and causes the weights to have a completely different meaning.

Definition 4. *Let w be a weighting vector of dimension n; then, a mapping $\mathbb{R}^n \to \mathbb{R}$ is an Ordered Weighted Averaging (OWA) operator of dimension n if $OWA_w(a_1, ..., a_n) = \sum_{i=1}^{n} w_i a_{\sigma(i)}$, where $\sigma(1), ..., \sigma(n)$ is a permutation of $1, ..., n$ such that the arguments are decreasingly ordered, i.e. $a_{\sigma(i-1)} \geq a_{\sigma(i)}$ for all $i = 2, ..., n$ (i.e., $a_{\sigma(i)}$ is the ith largest element in the collection $a_1, ..., a_n$).*

With this definition, weights are assigned to the position of the values rather than to the values themselves. Therefore, one may define different aggregation policies that give different importance to the highest or lowest values that have to be aggregated. In fact, the weighting vector of the OWA operator allows to move continuously from the minimum (when $w_n = 1$ and the rest are 0) to the maximum type of aggregation (when $w_1 = 1$ and the rest are 0). The compensative behaviour of the aggregation operator can be fixed by the set of weights. Compensation is the property

that a high degree of satisfaction in one criterion compensates a low degree of satisfaction in other criteria. The maximum operation (high orness) means full compensation or simultaneity (pessimistic aggregation policy), while the minimum operation (high andness) means no compensation or replaceability (optimistic aggregation policy). Those characteristics are especially suitable to combine the user's preferences in decision making processes and recommender systems.

In order to classify these OWA operators in relation to their conjunctive/disjunctive degree, a measure of *orness* α may be calculated for any weighting vector w of dimension n with Eq.2.1. The range of α is [0,1]. When orness is near 1 the weights define a disjunctive behavior, while an orness close to 0 means that the aggregation is conjunctive (low orness implies high andness, since these two measures are complementary).

$$\alpha(w) = \sum_{j=1}^{n} w_j \left(\frac{n-j}{n-1} \right) \tag{2.1}$$

Another characterizing measure of OWA weights is the *divergence*, which is a number in the range [0, 0.5]. The maximum divergence, 0.5, corresponds to the case of arithmetic average (i.e. equal weight for all the input arguments). The minimum divergence, 0, happens when only one input value is used (when $w_j = 1$ for a unique position j). Divergence reduces if the weights are assigned to a small subset of consecutive values.

$$div(w) = \sum_{j=1}^{n} w_j \left(\frac{n-j}{n-1} - \alpha(w) \right)^2 \tag{2.2}$$

Later, in 1997, Torra proposed the Weighted OWA (WOWA), which combines the OWA operator and the weighted mean WM [88]. The WOWA operator was introduced to model situations in which both the importance of the information sources and the aggregation policy have to be considered. The operator aggregates a set of values using two weighting vectors: one corresponding to the vector p in the weighted mean and the other corresponding to w in the OWA operator. The WOWA operator is defined as follows.

Definition 5. *Let p and and w be two weighting vectors of dimension n; then, a mapping $WOWA : \mathbb{R}^n \to \mathbb{R}$ is a Weighted Ordered Weighted*

Averaging (WOWA) operator of dimension n if

$$WOWA_{p,w}(a_1, ..., a_n) = \sum_{i=1}^{n} \omega_i a_{\sigma(i)}, \qquad (2.3)$$

where σ is defined as in the case of OWA (i.e., $a_{\sigma(i)}$ is the i-th largest element in the collection $(a_1, ..., a_n)$), and the weight ω_i is defined as

$$\omega_i = w^*(\sum_{j \leq i} p_{\sigma(j)}) - w^*(\sum_{j < i} p_{\sigma(j)}) \qquad (2.4)$$

with w^* being a monotone increasing function that interpolates the points $\left(\frac{i}{n}, \sum_{j \leq i} w_j\right)$ together with point (0,0). w^* is required to be a straight line when the points can be interpolated in this way.

2.4 Inference of missing preferences

In this section we present the procedure proposed to estimate the missing score for any leaf c of the ontology. This method has 3 steps:

Step 1. Find relatives. We find concepts that are semantically similar to c using the taxonomical relations of the ontology. Since only leaves have an associated TIS, we only retrieve concepts that do not have descendants. A set of related concepts is built by following the taxonomical relations in the ontology using Algorithm 1, where n is the number of similar concepts we want to find. The function *fathers* receives a set of concepts and an ontology, and it returns the set of direct ancestors of all the concepts in the input set according to the ontology. The function *leaves* receives a concept and an ontology, and it returns the set of ontology leaves that have a known TIS and belong to the subtree whose root is the given input concept. In the union operations, no repeated elements are stored in the output set.

Algorithm 1 Find Relatives

    Inputs: *concept c, user profile ontology θ, int n*

    Output: set of neighbor concepts.

1:   $F = $ fathers $(\{c\}, \theta)$

2:   $R = $ empty_set

3:   $R = R \cup$ leaves (f_i, θ) for all $f_i \in F$

4:   $m = |R|$

5:   **while** (m < n) and (F $\neq \emptyset$) **do**

6:      $F' = \cup$ (fathers $(f_i), \theta$) for all $f_i \in F$

7:      $R = R \cup$ leaves (f_i, θ) for all $f_i \in F'$

8:      $m = |R|$

9:      $F = F'$

10:   **end while**

11:   return R

In this algorithm we start searching for leaf concepts that are descendants of the father(s) of c, which can be found at different depths. If the number of elements is below the given input value n, we move to upper levels of the ontology to find leaf concepts descending from the hierarchy with root in a grandparent of c. Iteratively, if the number of neighbors with known score is still low, we continue exploring other regions of the ontology by going upwards in the chain of ancestors of the first concept c.

Step 2. Concept Importance. The determination of the importance of each relative is done according to its semantic similarity to the given concept c. As explained in section 2.3.1, there is a large set of semantic similarity measures available in the literature. The most appropriate measure depends on the purpose of each problem. In the formulation presented in Eq. 2.5, the semantic distance *dsem* is not specified.

Despite any semantic similarity could be used, we suggest the use of Path Length. A distance based on steps is appropriate taking into account that c will be a very specific concept (located in the leaf of a branch). The more steps up and down are needed to find another leaf concept, the lower is their degree of semantic similarity.

Weights are defined in Eq. 2.5. The idea is that we want to give more importance to the concepts that are close to the target concept c, because

2.4. Inference of missing preferences 23

they represent tags with a strong semantic similarity to c and, thus, their interest scores are expected to be similar to the value that we have to estimate for c. For instance, concepts at distance 2 (i.e. brothers) are given more relevance than concepts at distance 3 (which are uncles or nephews). Concretely, for a concept r_k found at distance d with respect to the target concept c, the corresponding weight p_k is calculated with the following expression, in which D is the maximum distance at which a related concept has been retrieved, and $\#concepts(d)$ is the number of related concepts found at a certain distance d:

$$ p_k = \frac{1}{\Omega \cdot dsem(c, r_k)}, \text{ where } \Omega = \sum_{d=2...D} \frac{\#concepts(d)}{d} \tag{2.5} $$

Step 3. TIS calculation. The estimation of the tag interest score of c using the set of relatives R and the weighting vector p is done by means of the aggregation of the known scores of these relatives. The value of *TIS(c)* is calculated using the WOWA averaging operator on the known scores of the relatives of c. As explained before, the classical OWA weights allow the definition of different aggregation policies. With conjunctive parameters, the resulting score is penalized when similar concepts have low scores (pessimistic approach), whereas with disjunctive parameters the score is based only on the highest scores of the similar concepts (optimistic approach). A neutral configuration is also possible, which leads to the classic arithmetic average.

In order to apply the operator, first the aggregation policy must be specified by defining a weighting vector w of size $|R|$. This vector can be manually defined by the user or it can be automatically constructed. Yager described that the weights can be obtained with Eq. 2.6 using a linguistic quantifier, which is a function Q that is defined according to the quantity of simultaneous values to take into account (e.g. "most", "at least half" or "all") [97].

$$ w_j = Q\left(\frac{j}{n}\right) - Q\left(\frac{j-1}{n}\right) \tag{2.6} $$

Different linguistic quantifier functions may be obtained by setting a certain degree of orness. For example, $Q(r) = r^\alpha$ [97].
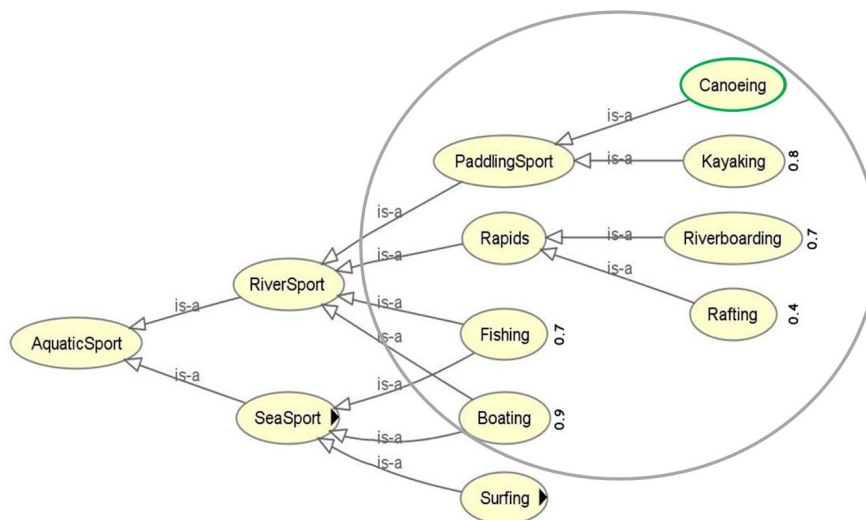
Figure 2.2: Set of relatives and TIS used for the calculation of the new tag interest score for Canoeing.

To summarize, the proposed algorithm to estimate the missing preference score of a concept c is the following:

Algorithm 2 TIS calculation

   Inputs: concept c, user profile ontology θ, int n, OWA weights vector w

   Output: score (TIS) for concept c

 1: $R =$ find relatives (c, θ, n) with known TIS
 2: sim= calculate weights with semantic distance (R, θ, c)
 3: score= $WOWA_{sim,w}$ (TIS of concepts in R)
 4: return score

Example: Let us consider that Mr. Smith has the user profile shown in Fig. 2.2.

We want to calculate the new TIS for the concept "Canoeing", which is unknown at the moment. We will use the information of the neighbor concepts following Algorithm 2, as explained. First, the aggregation policy of OWA must be chosen. In this example, a conjunctive model with small

orness will be used, so we take $\alpha = 0.25$. The weights w_k will be assigned later, depending on the number of relatives found in the ontology. In Step 1, we use the method Find Relatives (Alg. 1) and we get 5 relatives (grey area in Fig. 2.2): R = {Kayaking (TIS=0.8), Fishing (TIS=0.7), Boating (TIS=0.9), Riverboarding (TIS=0.7) and Rafting (TIS=0.4)}. As we know now that there will be 5 input arguments, we can establish the OWA weights:

$$w = (0.0, 0.0, 0.33, 0.33, 0.33)$$

In Step 2, we calculate the weight of each concept using path length as the distance with respect to "Canoeing". These distances are the following: Kayaking $dsem=2$, Boating $dsem=3$, Fishing $dsem=3$, Riverboarding $dsem=4$, Rafting $dsem=4$. Notice that Kayaking is a brother concept (smallest distance), while Riverboarding and Rafting are the less similar. The largest distance in this case is 4. Thus, using Eq. 2.5 we get $\Omega = 1/2 + 2/3 + 2/4 = 1.66$ and

$$p = (0.3, 0.2, 0.2, 0.15, 0.15).$$

In Step 3, the relatives are ordered in a descending way depending on their TIS. The three with less TIS will be used for the estimation of the interest on Canoeing. As shown in w, the three of them will have the same contribution. Their weights on the final calculation depend on the semantic distance, being Rafting and Riverboarding less influent than Fishing. The WOWA operator can be applied with these input values:

$$WOWA_{(0.3,0.2,0.2,0.15,0.15),(0.0,0.0,0.0,0.33,0.33,0.33)}(0.8, 0.7, 0.9, 0.7, 0.4) = 0.61$$

2.5 Experiments

In order to validate the method for inferring missing scores that were not provided by the user explicitly, an experimentation procedure has been defined to perform multiple tests with different configurations. Several user profiles have been manually defined in order to deal with different situations, so they do not correspond to real people. The testing procedure is as follows:

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

Chapter 2. Representing and inferring preferences in
26 **ontology-based recommender systems**

1. Take a predefined user profile ontology that has a TIS for all the leaf concepts.

2. Remove a percentage of the TIS values randomly to simulate that the user has not entered some of the interest scores. These will be the missing values to estimate.

3. Use the estimation method based on WOWA to assign a TIS to each of the leafs without preference value.

4. Compare the original TIS with the calculated TIS.

5. Repeat steps 2-4 a certain number of times and calculate the average error.

Different parameters are used in this procedure. The repetition of the tests several times with different subsets of missing scores enables the calculation of a better quality indicator. The following subsections describe the experimental setting and the data used for the validation of the proposed method.

2.5.1 Experimental setting

The data used in the tests corresponds to people that is going to visit a touristic place in their holidays. Two ontologies that describe different types of activities (Leisure and Sports) have been used. The ontology-based user profiles used in this case study have different ratios between the number of concepts that the user considers interesting (*likes*) and the number of concepts that the tourist is not interested in (*dislikes*). Some profiles correspond to tourists that are interested in a large variety of activities, while others search for a very specific type of touristic attractions. This will enable us to study the behavior of the proposed method to estimate the missing scores in different situations. After presenting the ontologies and profiles, the parameters used in the automatic testing procedure are given.

2.5.   Experiments   27

Ontologies and user profiles

- Leisure Ontology (Figure 2.3): it distinguishes 3 classes in the most
  general level: City Activities (Day Life and Night Life activities),
  Relaxation (Beach Activities, Spa and Wellness activities) and Amuse-
  ment Parks (Natural Parks and Theme Parks). This ontology has
  40 leafs (basic concepts). Its maximum depth is 6 and the average
  branching factor is between 2 and 3.

  ○ Leisure Ontology - General Profile (L1): likes = 27, neutral
    = 4, dislikes = 9. This user prefers relaxation activities, espe-
    cially beach walking, beach picnic, body care, massages, yoga,
    whirlpool bath and jacuzzi. He/she also likes amusement parks
    and day life city activities like sightseeing, gastronomy fairs and
    craft market. On the other hand, he/she dislikes music activities
    like concerts or discos, as well as game-related activities.

  ○ Leisure Ontology - Specific Profile (L2): likes = 9, neutral = 4,
    dislikes= 27. This case corresponds to a family with children
    that makes a visit for a weekend. This family is looking for
    amusement parks (water park, aquarium or jungle trek), and
    they also are interested on beach activities. This family does not
    want to do gastronomy-related activities, relaxation activities,
    botanical activities or shopping.

  ○ Leisure Ontology - Balanced Profile (L3): likes = 20, neutral
    = 3, dislikes = 17. This user has a similar number of likes
    and dislikes. The most preferred activities are sightseeing, craft
    market, gastronomy routes, typical food or national park visits.
    He/she is not interested in jungle trek parks, water parks, and
    relaxation or care activities.

- Sport Ontology (Figure 2.4): it divides sports in 3 main classes:
  Land Sports (sports in the forest, on the mountain, motor sports and
  shooting activities), Air Sports (gliding, parachuting, aerobatics and
  balloon activities), and Aquatic Sports (sea sports and river sports).
  This ontology has 58 leaves. Its maximum depth is 7 and the average
  branching factor is between 3 and 4.

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

Chapter 2. Representing and inferring preferences in
28   ontology-based recommender systems
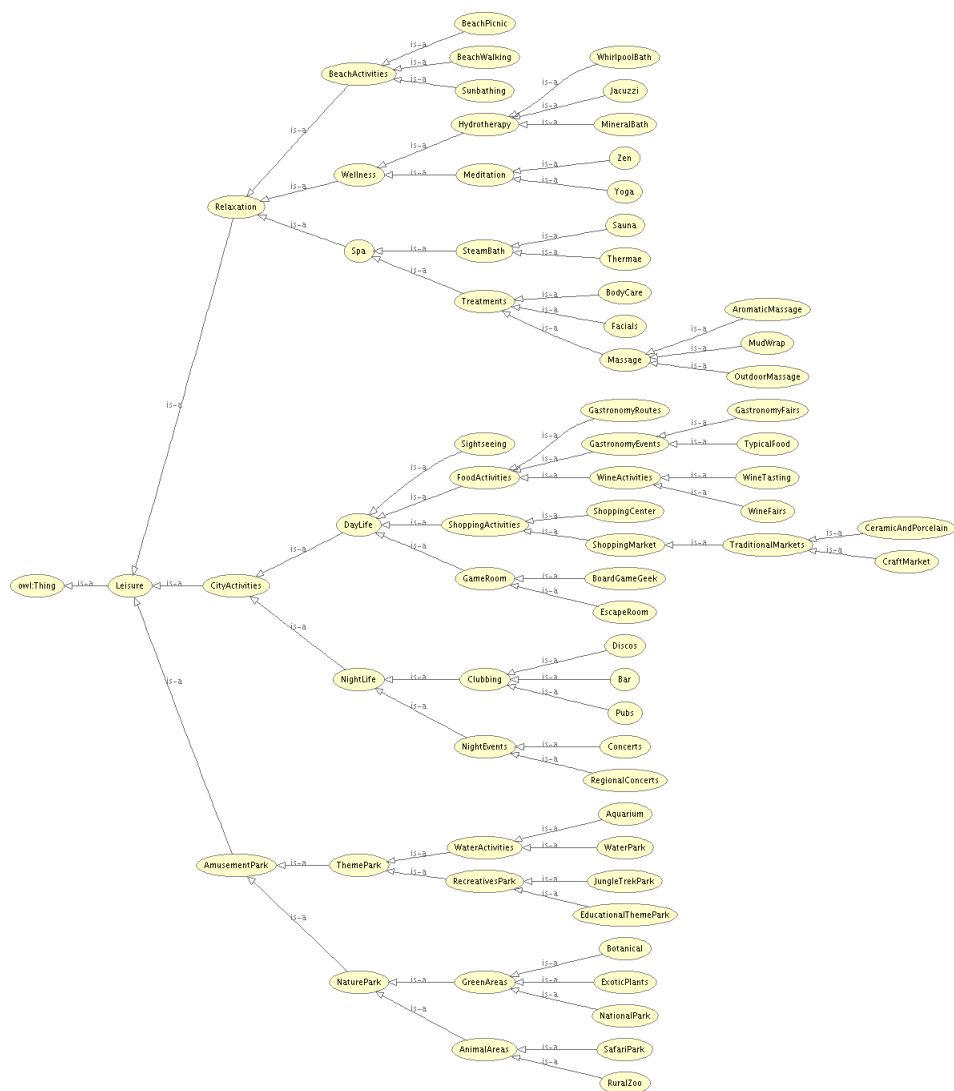
Figure 2.3: Leisure Ontology
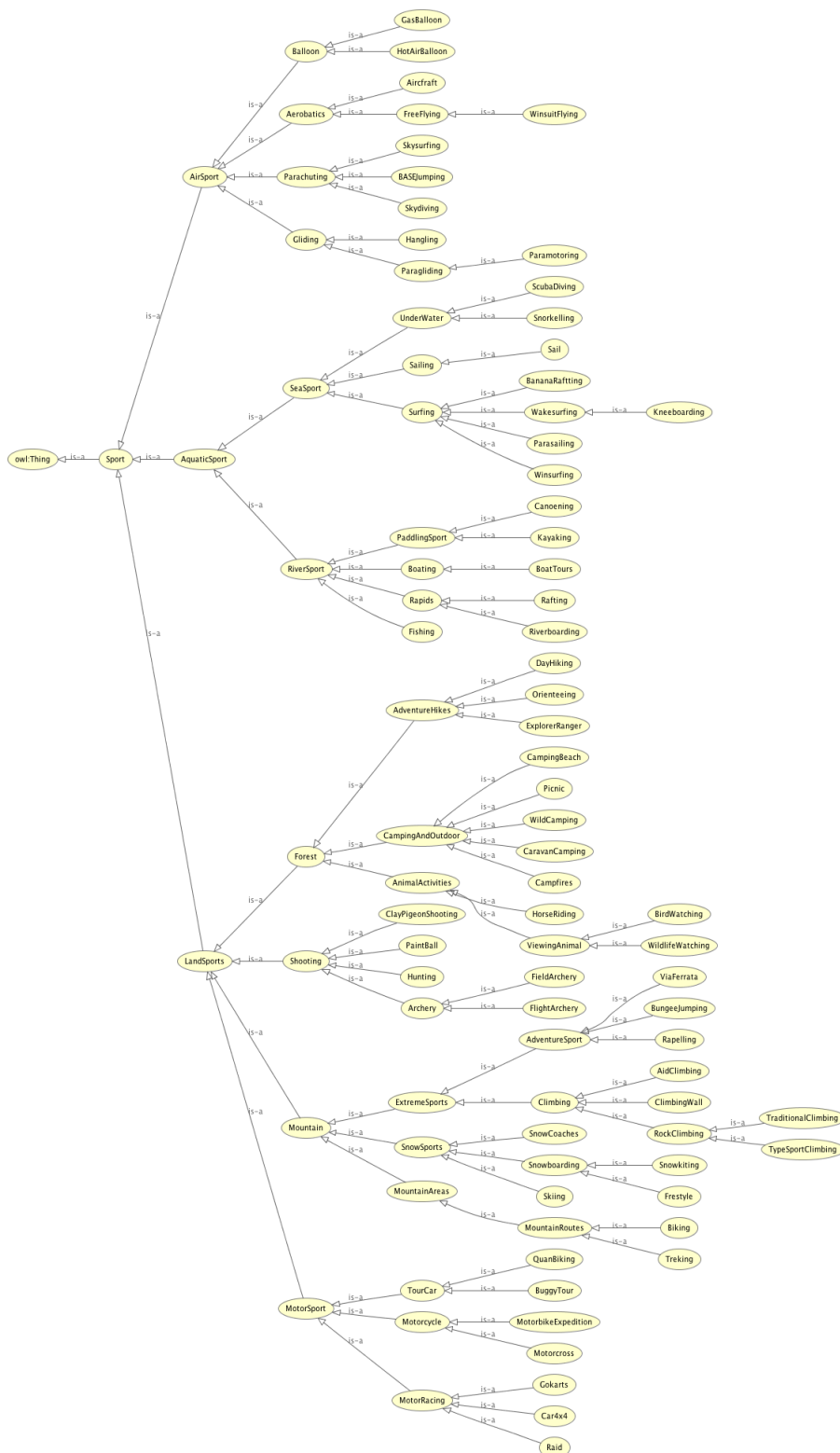
Figure 2.4: Sport Ontology

○ Sport Ontology - General Profile (S1): likes = 40, neutral = 3, dislikes = 15. This is a sportive tourist who is eager to practice most kinds of sports. He/she only dislikes exploring and camping activities, archery and fishing.

○ Sport Ontology - Specific Profile (S2): likes = 15, neutral = 3, dislikes = 40. This user prefers mountaineering and river activities like trekking, wall climbing, rafting or canoeing. In the trip, he/she wants to avoid biking, picnic, horse riding and motor activities, among others.

Parameters used in the test

To run the experiments the following values have been used for the different parameters:

- Percentage of missing preference values: from 5% to 50%, in steps of 5%.

- Minimum number of relatives: 2, 4, 6 and 8.

- OWA aggregation policy with a divergence of 0.025 (which corresponds to the use of approximately half of the values) and with two degrees of orness:

    ○ Pessimistic aggregation with $\alpha = 0.2$ (conjunctive).

    ○ Optimistic aggregation with $\alpha = 0.8$ (disjunctive).

- Number of repetitions = 20 times.

2.5.2   Validation index

The quality of the new interest scores is measured as the *root-mean-square error* (RMSE) between the predicted scores and the original ones. As tags without score are selected randomly, each test has been repeated 20 times and the average and deviation of the RMSE have been calculated.

RMSE is a common validation index to measure the differences between the observed population values, $\hat{y}_i$, and the values predicted by a model, y_i. These individual differences are called *residuals* when the calculations

are performed over the data set that was used for estimation, and they are called *prediction errors* when computed on new data (which is our case).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \qquad (2.7)$$

2.5.3   Results

In this section the obtained results are shown separately for each user profile. In particular, we focus on identifying the best number of relatives to use in pessimistic and optimistic aggregation policies. It is also interesting to know if the number of missing scores has any influence on the number of relatives needed for the estimation. After analyzing each user profile, we try to identify common guidelines that could be used to decide when to use the optimistic or the pessimistic approach, as well as the number of relatives to consider.

Analysis of the RMSE in different user profiles

For each profile, two figures are presented: on the left, the RMSE obtained with the optimistic policy, and on the right, the RMSE with the pessimistic policy. The horizontal axis shows the different proportions of missing values studied, from a case where the profile is almost complete (only 5% missing scores) to a profile with just half of the possible interest scores available (50% missing TIS). Each line represents the results with a different number of minimum related concepts (even numbers from 2 to 8).

Profile Leisure Ontology - General Profile (L1)
In this first test, results are quite different for the optimistic and pessimistic types of aggregation. In the optimistic case, we can see that the best result (lowest RMSE) is obtained with 8 concepts, except when the number of missing scores is below 10%, where it is enough to use 2 or 4 concepts. This is probably because the ontology is full of TIS and, hence, we have a good knowledge of the users preferences and we do not need much additional evidence to predict a correct value for the missing scores. On the contrary, when using a pessimistic or conjunctive aggregation operator,

we need a majority of concepts in agreement in order to assign a high score, therefore it is more probable to predict low scores, which is not appropriate for this tourist, because he/she is interested in many different types of activities. Consequently, in this case, it is better to use only 2 relatives for the conjunctive aggregation. Furthermore, we can see that the error is lower in the optimistic setting (under 0.3 in most cases) while it is above 0.31 in most of the cases of the pessimistic approach, even with 2 relatives. Variance is similar for the different percentages of missing scores, but it is a bit larger for the pessimistic case.
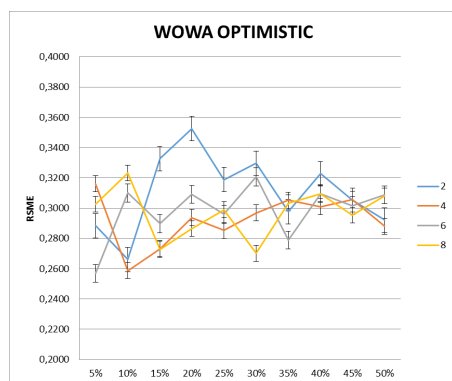
Figure 2.5: RMSE with an optimistic WOWA in profile L1
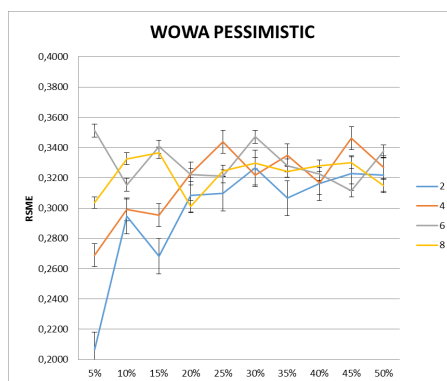
Figure 2.6: RMSE with a pessimistic WOWA in profile L1

Profile Leisure Ontology- Specific Profile (L2)

In this case the user has a small number of preferred activities. The worst error in the optimistic case (Figure 2.7) ranges from 0.27 to 0.29 and it is generally obtained with 2 concepts. It can be observed that, in this case, it is better to use 6 or 8 relatives in the optimistic approach. On the other figure (Figure 2.8) the conclusions are a bit different. It corresponds to the pessimistic (i.e. conjunctive) approach, where the minimum error is obtained with 2 or 4 neighbors (with a very small difference). In this case, 8 relatives give the worst RMSE. These results are similar to those obtained with user L1, in spite of the differences in the balance between likes and dislikes in L1 and L2.

Regarding the best RMSE levels, we can see that the minimum error is around 0.24 in both cases (optimistic and pessimistic). In Figure 2.8,

the error changes depending on the amount of scores available in the user profile ontology. When the knowledge is large (5-20% of missing data values), the error is smaller than the situations where the ontology has less information (above 25% of missing values). However, in the optimistic case, this difference is not appreciated. It can be seen in that in the pessimistic case the error variance is larger than in the optimistic case, in which it is more stable.
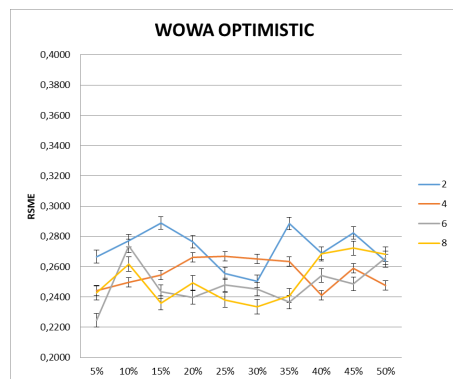
Figure 2.7: RMSE with an optimistic WOWA in profile L2
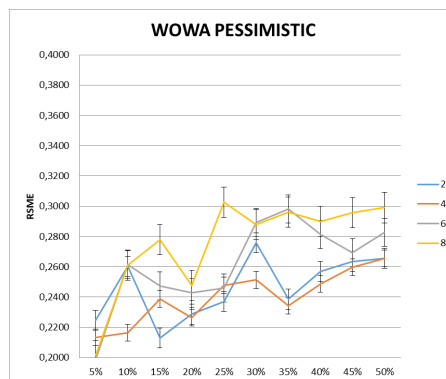
Figure 2.8: RMSE with a pessimistic WOWA in profile L2

Profile Leisure Ontology- Balanced Profile (L3)

In the Leisure ontology we tested a third type of user. In this user profile, L3, half of the tags are positively scored and the other half are negatively scored.

In Figures 2.9 and 2.10 we can see that both versions (pessimistic and optimistic) get the best RMSE with 6 and 8 tags, with the best values between 0.30 and 0.32. In both cases, when the number of missing scores is large (above 40%), the results with 6 neighbors are a little bit better, although the difference is small. When 2 neighbours are considered the error is larger in the optimistic aggregation. This is not the case for 6 or 8 relatives. In both cases the RMSE is around 0.32.

Profile Sport Ontology-General Profile (S1)

This profile corresponds to the Sport semantic criterion, which has a different ontology, as described above. The first test with the Sport ontology corresponds to a tourist that is very keen on doing different types
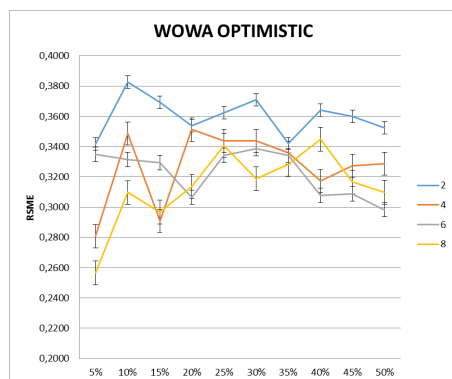
Figure 2.9: RMSE with an optimistic WOWA in profile L3
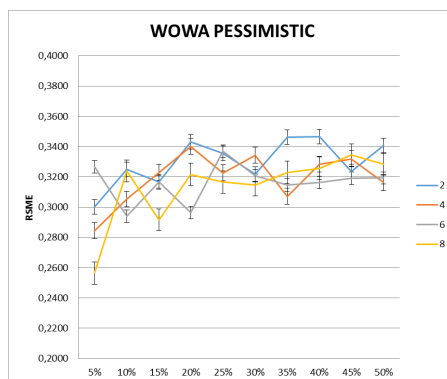
Figure 2.10: RMSE with a pessimistic WOWA in profile L3

of sport.
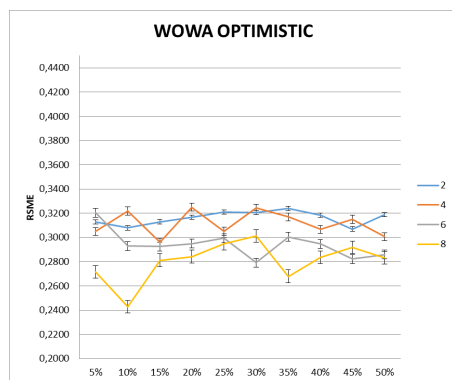
Figure 2.11: RMSE with an optimistic WOWA in profile S1
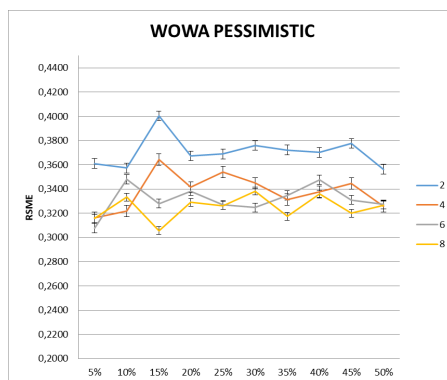
Figure 2.12: RMSE with a pessimistic WOWA in profile S1

The distribution of the taxonomical relations in the Sport ontology leads to different RMSE values. In Figure 2.11 and Figure 2.12 we can observe more stable and differentiated RMSE lines for each different number of relatives. Using only 2 values is the worst option, while using 8 is generally the best. In this case, with a user with a large number of concepts with high interest (TIS>0.5), we can see that the optimistic approach leads to a lower error, oscilating between 0.27 and 0.32. The conjunctive approach,

which is more conservative, obtains errors between 0.31 and 0.40, clearly higher than the optimistic one.

Profile Sport Ontology- Specific Profile (S2)
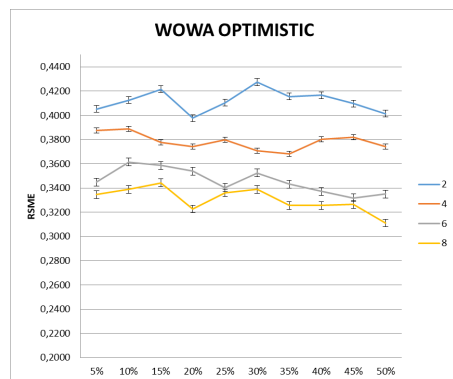
Figure 2.13: RMSE with an optimistic WOWA in profile S2
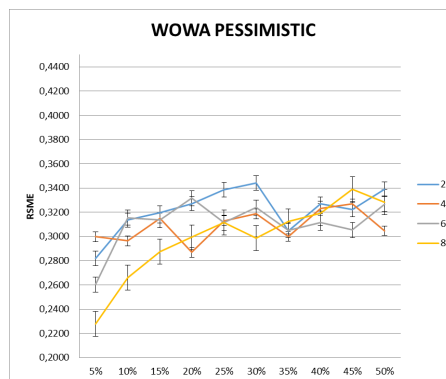
Figure 2.14: RMSE with a pessimistic WOWA in profile S2

The RMSE graphical lines are again quite stable, showing more clearly the difference in the error depending on the number of neighbors used for the prediction of the missing value. This is clearer in the optimistic approach (Figure 2.13) than in the pessimistic one (Figure 2.14). With more tags we reduce the error to values between 0.32 and 0.34 using the optimistic aggregation. Taking into account that this profile corresponds to a tourist searching for specific sports (likes=15), the error made with a conjunctive approach is smaller (with RMSE close to 0.3). In this pessimistic case, when the amount of missing data is large (above 30%), it is better to use 4 or 6 neigbours rather than 8, because they will be widespread in the ontology and they will be related to very different kinds of sports.

After this study of the five user profiles, four main conclusions are drawn:

- When the user is searching for specific tags, WOWA should use a pessimistic policy.

- In the pessimistic model, the number of relatives should be low when the percentage of missing scores is above 35% (to perform a local

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

Chapter 2. Representing and inferring preferences in ontology-based recommender systems

36

focused search). In the case of an almost complete profile, we can increase the number of relatives to be used in order to improve the prediction.

- When the user has a profile with more likes than dislikes, WOWA should be optimistic and the use of more relatives is recommended (i.e. aggregate the information of 8 related concepts).

- The error may be different depending on the amount of missing scores in the user profile ontology. It is appreciated a difference between profiles with more or less than 20% of concepts without known TIS.

Analysis of the number of concepts used for the calculation of a tag interest score

In order to study in more detail the influence of the concepts used for the calculation of the unknown scores we have analyzed the number of relatives used in each calculation. The following bar charts show the averaged percentage of times that a certain number of relatives has been used during the 20 tests. Figures 2.15, 2.16, 2.17 corresponds to the tests with the Leisure ontology and Figures 2.18, 2.19, 2.20 to the ones with the Sport ontology. Each figure displays the bar chart of 3 situations (with 10%, 30% and 50% of missing data). Each bar corresponds to the given number of minimum relatives to retrieve: 2, 4, 6, and 8.

We can observe that, even though a minimum number of tags to use has been fixed, depending on the distribution of the TIS in the ontology the algorithm needs to go upwards in the taxonomy and consider sometimes many different branches. Therefore, the actual number of known scores may be larger than the minimum required.

Figures 2.15, 2.16, 2.17 show the histograms in percentages of the 3 different situations, from the best case (when we know most of the users preferences) to the worst (with just half of the information). Analyzing the 3 situations, the following facts can be observed:

- 10% of missing data: indicates a situation with a lot of known information about the user (number of TIS available is high = 35). In this graphic, we can see that when fixing 2 or 4 neighbors, there is a high percentage of times that less than 5 concepts are used. When

2.5. Experiments   37
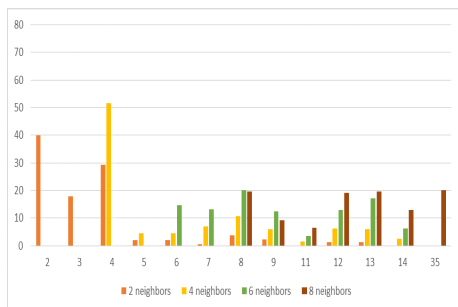
Figure 2.15: Missing values: 10% of 40=4 and TIS available: 35
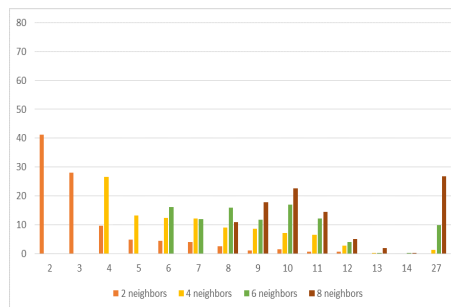
Figure 2.16: Missing values 30% of 40 = 12 and TIS available: 27

6 or 8 concepts are the minimum, the distribution is quite stable until 14 concepts. It is worth noting that the whole set of available concepts (i.e. 35) is used in 20% of the WOWA calculations when we set a minimum of 8 neighbors. This situation corresponds to the case where the algorithm has to search in the whole ontology.

- 30% of missing data: indicates a situation in which we have 1 unknown preference for each 2 known TIS. In this case, the maximum number of available scores is 27 (see horizontal axis in the central graphic). 25% of the times in which 8 relatives were needed required the use of the whole set of tags. Again, fixing a lower number of relatives is directly related to using less concepts, especially for the cases of
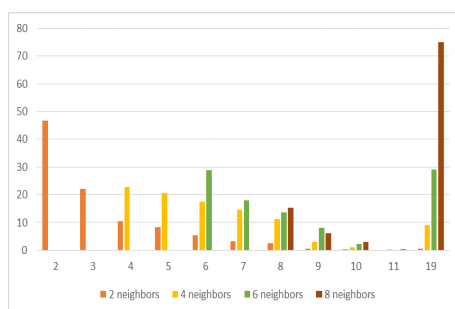
Figure 2.17: Missing values 50% of 40 = 20 and TIS available: 19

2 and 4. We observe a gap between 14 and 27 concepts, which is directly related to the number of children of the root of the ontology.

- 50% of missing data: it is the worst case, corresponding to initial stages of the recommendation process, when the user has only introduced half of the tag interest scores (low number of TIS available). In this case, the proportion of used data is significantly higher in the last bar (when using all the 19 tags), especially with 8 neighbors, but also with 6.
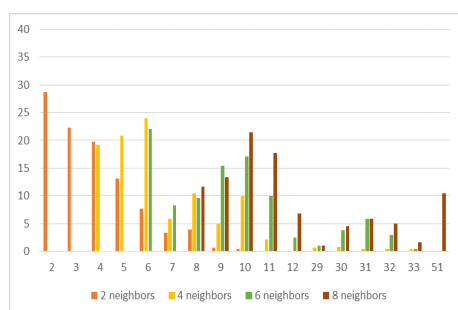
Figure 2.18: Missing values: 10% of 60 =6 and Tags available: 51
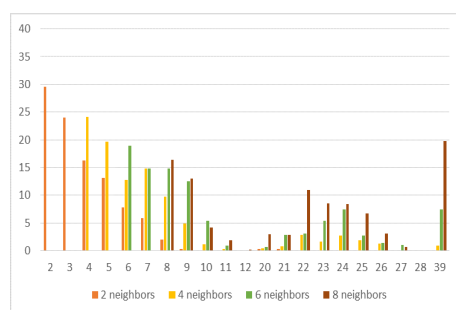
Figure 2.19: Missing values 30% of 60 = 18 and Tags available: 39
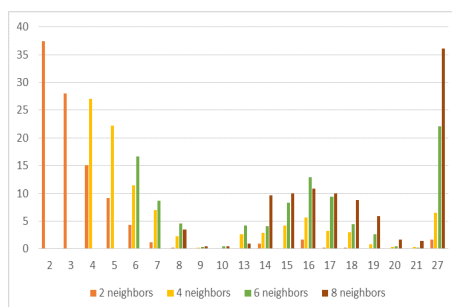
Figure 2.20: Missing values 50% of 60 = 30 and Tags available: 27

Likewise, in the case of the sports activities we also want to know the percentage of tags used in the three different situations (Figures 2.18, 2.19, 2.20). The observations are analogous to the case of the Leisure ontology:

- 10% of unknown values: when the number of TIS available is high (51), we observe that the proportion of the concepts used when fixing 6 and 8 neighbors is similar, where the actual number of tags used is usually between 8 and 11 concepts. The whole set is only used 10% of times when 8 relatives are required.

- 30% of unknown values: indicates a situation in which we have 1 unknown preference for each 2 known TIS, as before. In this case, the bars are higher on the left of the graphic (low number of used tags). They decrease until 12 tags and then increase again to 22-25. A second gap without bars happens between 28 and 39. This is again related to the number of steps upwards that the algorithm must do until it finds the required number of concepts. The two first levels of the ontology have a great influence in these numbers because they form subgroups of related concepts.

- 50% of unknown values: it is the worst case, corresponding to initial stages of the recommendation process. In this graphic we note a large percentage of cases concentrated in the first bars. The distribution of the concepts in the different semantic subgroups modelled in the ontology makes that certain number of tags are not found (9-10 and 20-26) in this case.

Comparison with other methods

Table 2.1 showed a list of methods that deal with user profiles stored in ontologies. In order to compare the proposed WOWA-based preference inference method with the previous ones, first we selected three methods that work in the same conditions: the user introduces a first subset of numerical scores on some of the most specific concepts, no other sources of knowledge are used, and there isn't any interaction with the user. The selected methods are labelled as M1 [81], M2 [16] and M3 [76]. In all of them there is some kind of spreading (through an ontological structure) of a subset of initial preference scores given by the user.

In the work of [81] (M1) a weighted average is also used to estimate a missing score. A set of close concepts is taken from the ontology. The weights considered in the averaging procedure are said to be based on the number of items under each concept, but, as no more details are given,

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

Chapter 2. Representing and inferring preferences in
40   **ontology-based recommender systems**

they are not replicable. Therefore, we will use the same weights proposed in Eq. 2.5. The score of a concept is calculated from the scores of similar concepts.

Method M2 [16] also proposes a similar averaging procedure based on a set of weights that are interpreted as the probability that one concept is relevant to estimate the score of the other. The authors only indicate that the definition of the weights is critical and very hard to decide. In their experiments, weights were empirically fixed. The new score is obtained with a spreading activation mechanism over the semantic network. We have replicated this spreading activation to the leaves that do not have any score. The weight of the connection between concepts has been calculated as proposed in Eq. 2.5 in order to compare the effect of the activation process. Concepts with a missing score are set to zero, and then they are activated by performing an aggregation of the scores of the neighbor concepts. These concepts are ordered decreasingly by its interest score in vector X. Then, the following equation is used:

$$score(c_j) = score(c_j) + (1 - score(c_j)) * w_i * score(x_i) \ , \ for \ all \ x_i \in X$$
$$(2.8)$$

The initialization method M3 [76] consists in calculating a weighted average of the interest scores of the neighbor concepts using as weight a semantic similarity measure based on finding the Least Common Ancestor (LCA) and the distance of each neighbor (c and d) to the LCA (Eq. 2.9).

$$dsim(c, d) = \rho * \frac{|ancestors(LCA(c, d))|}{|ancestors(c)|} + (1 - \rho) * \frac{|ancestors(LCA(c, d))|}{|ancestors(d)|}$$
$$(2.9)$$

In order to make these three methods comparable to the one proposed in this paper, we consider the same set of neighbors in all the cases. Neighbors are found as proposed before in Algorithm 1. Therefore, what we change is the way of aggregating the contribution of each of those similar concepts. We will use the ontology of Sports and the two user profiles S1 (general) and S2 (specific). As the best performance was obtained with 6 and 8 neighbors, we have fixed the number of neighbors to 7 for all the methods.

In Figures 2.21 and 2.22 we can see the performance of the 3 methods found in the literature and the new WOWA-based one. An optimistic

WOWA was chosen for the most general profile (S1), whereas the pessimistic one was applied in the case of the specific profile S2, as they have been previously shown to be the most appropriate for each case. It can be seen that WOWA outperforms methods M1 and M2 in almost all the cases. The exception of M1 in S1 with 45% of missing values may be due to the randomness in the selection of the tags in the tests. Method M3 gives a performance comparable to WOWA in the S1 profile, being sometimes 0.02 points better or worse in RMSE. However, in profile S2, despite the fluctuations of the WOWA method, it is always the best one. We can notice that, in the case of profiles with more than 45% of missing information, WOWA is outperformed by M3 in both cases.
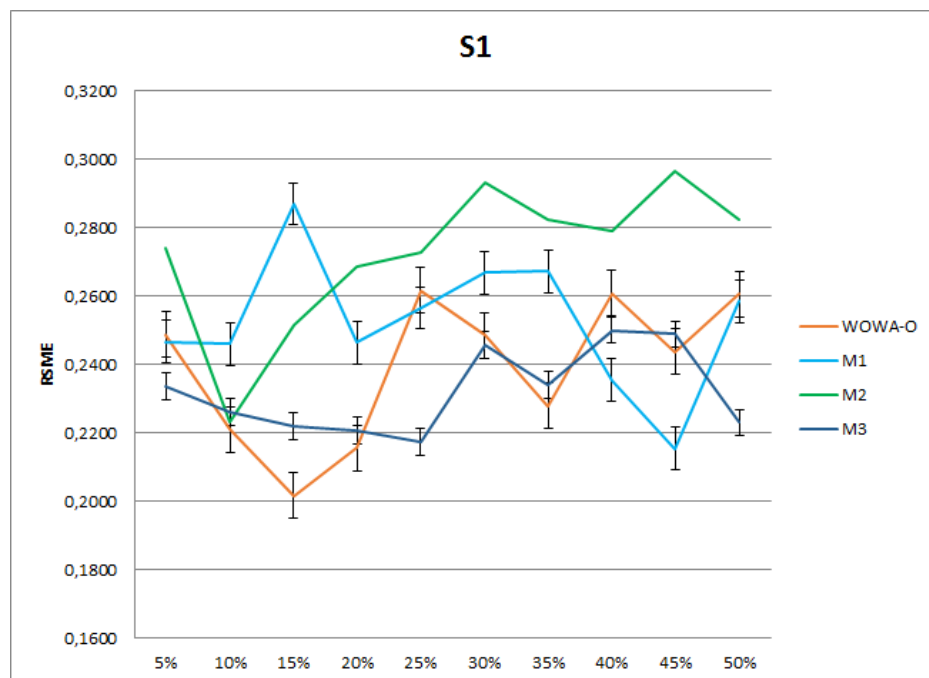
Figure 2.21: Comparison of RMSE in profile S1

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

Chapter 2.   Representing and inferring preferences in ontology-based recommender systems

42
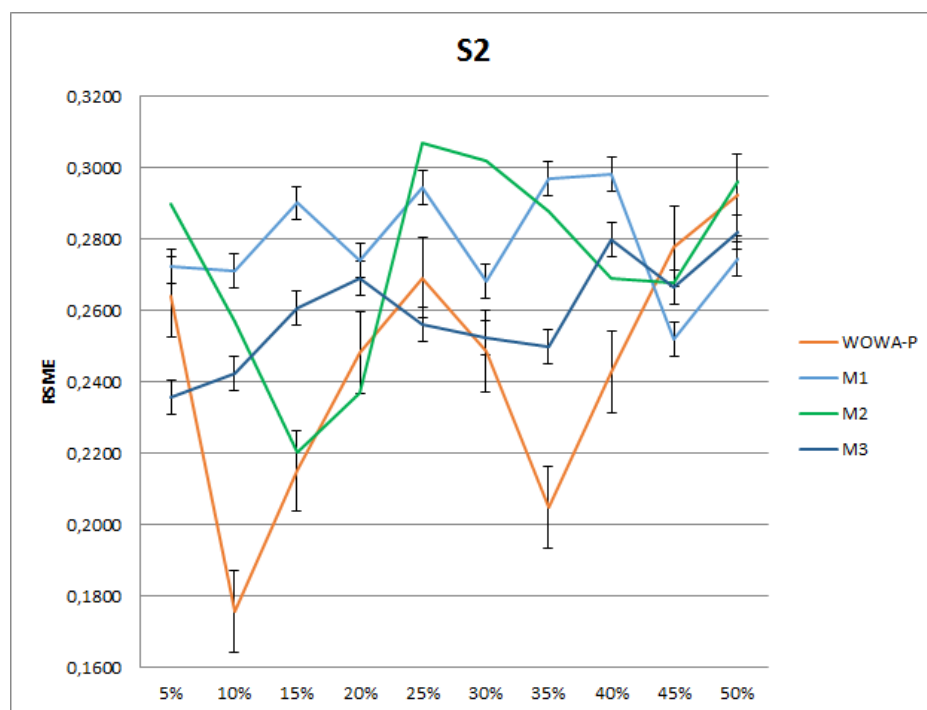
Figure 2.22: Comparison of RMSE in profile S2

2.6   Conclusion

In this chapter, the initialization of semantic user profiles has been studied. The chapter proposes the use of ontologies to represent the preferences of the user by means of a numerical indicator denoted as *Tag Interest Score* (TIS) associated to the set of leaves of the ontology. The main contribution is the formalization of a method for estimating the interest score of a concept using information of other TIS available in the ontology. The proposed procedure is based on the WOWA aggregation operator that enables the modeling of the aggregation using two sets of weights. By means of a semantic similarity measure, the importance of the concepts whose preferences are known is introduced to guide the aggregation. This is a novel procedure to automatically adjust the weights depending on the concept studied in each case, exploiting the structure of

the ontology, where multiple inheritance is possible.

In order to validate the method for inferring missing scores a case study of touristic recommender systems has been considered, using two ontologies and several profiles. Results showed that the parameters of the WOWA operator must be appropriately defined according to the type of user. It was also observed that it is generally recommended to use more information from related concepts, rather than just to focus on a specific neighborhood. However, the second study shows that when the knowledge about the user preferences is small, this may be problematic because many predictions will be done using all the scores of the ontology, which will lead to the same prediction for quite different concepts. The advantage of WOWA with respect to other methods in the literature is the ability of using the information provided by the neighbors in different ways so that it can be more optimistic or pessimistic in the predictions, while other approaches apply some kind of weighted average. Knowing the user's personality towards the recommended items, we can then adjust the WOWA parameters to better estimate the unknown preference scores. The new algorithm has also been compared with three methods of the state of the art, obtaining a comparable performance.

Once we have a method for constructing a semantic user profile, we will now study how to exploit this knowledge by using multiple-criteria decision aiding procedures in recommender systems.

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

<div style="text-align: right;">**3**</div>

Multiple Criteria Decision Aiding

Multi-Criteria Decision Aiding (MDCA) is a research field that has gained significant attention from researchers in the last years [28]. It started as a subdiscipline of Operation Research, but now it has strong links with many other disciplines, such as Computer Science in general and Artificial Intelligence in particular [56]. These methods have been widely applied to support decision makers in different fields of application, especially in environmental management and project management [34], [12].

The use of MCDA techniques in classic recommender systems has not been studied in depth until now. Although there have been some studies that have proposed the use of aggregation operators for merging the values of multiple criteria, these approaches are quite simple and do not take advantage of the well-established MCDA methods [37], [2], [95].

In this chapter, we will first focus on the presentation of the basic notions and methods of multi-criteria decision support. In the following chapter, we will propose a modification of the ELECTRE method that is able to take into account semantic criteria by means of ontology-based user profiles as defined in the previous chapter.

3.1 Basic elements in the MCDA model

The formalization of a multiple criteria decision aiding procedure is made from the point of view of the so-called *Decision Maker* (DM), who

is the person that must make a decision. The data required in a decision problem is represented by means of two key elements:

1. **Alternatives.** They are the potential actions for the decision problem from which the decision maker has to decide. Alternatives are represented as follows:
   $A = \{\ a, b, c, ... \ \}$ is the finite set of alternatives and n is the number of alternatives in A.

2. **Criteria.** A criterion g is a tool constructed for evaluating and comparing potential actions according to the DM subjective point of view about some reference indicators. This evaluation must take into account, for each action a, all the pertinent indicators (i.e. variables or attributes), and it must also consider the preferences of the DM. Criteria are represented as follows:
   $G = \{\ g_1,\ g_2,..,\ g_m\ \}$ is the finite set of criteria, in which m is the number of criteria in G. Then, $g_j(a)$ represents the performance value of alternative $a \in A$ on criterion $g_j \in G$.

3.1.1 Types of criteria

In classic MCDA approaches, the construction of criteria is usually done by an analyst with the DM through an interactive *decision process*. Taking part in a decision process as a MCDA analyst consists in helping the DM to define the decision problem by identifying which are the relevant variables having an influence on the decision. Only these variables must be included in the process.

In recommender systems, the set of variables is usually defined by the person that defines and constructs the system. In this case, this person may be considered as the analyst and decision maker at the same time. This person decides which will be the attributes that will be associated to each alternative in the system. These objective attributes have to be transformed into subjective suitability criteria. Another difference in recommender systems with respect to typical MCDA systems is that there is not a unique DM. On the contrary, the recommender system is expected to be used by many people with different preferences. Therefore, a database of user profiles is included in this kind of systems. Criteria, then, must

extract the information from the user profile in order to be able to evaluate the alternatives for each different user in a personalized way.

In the decision procedure, criteria are used to compare two actions according to their performance on a certain attribute. For a correct comparison, the meaning of the performance values must be considered. The elements $g_j(a)$ are called degrees or scores of the scale. It is important to distinguish two types of scales:

- Ordinal scale: they just indicate an order between the scores. Thus, the gap between two consecutive degrees does not reflect the strength of the difference of preference between them. It can be linguistic or numerical.

- Quantitative scale: scale whose degrees are defined by referring to a clear concrete defined quantity. In that case the difference between two values indicates the quantity of preference difference.

Depending on the type of scale, some mathematical operations may not be appropriate. In this work, we will concentrate on the use of the Tag Interest Score (TIS) defined in the previous chapter. Therefore, we will use a quantitative scale of measurement for semantic variables.

3.2 Types of problems

The type of MCDA problem depends on the goal of the decision maker when approaching a certain problem [28]. Three main types are distinguished:

- Choice problem: it is a selection problem where the DM has to choose the best option. MCDA helps the DM to find among all the possibilities the best alternative(s).

- Ranking problem: in this type of problem, the DM wants to get a ranking of the alternatives from the best to the worst. Two (or more) alternatives may be equivalent. The outcome of a MCDA method can be either a partial or a complete ranking of the set of alternatives.

- Sorting problem: the problem of sorting consists in assigning the alternatives to predefined categories. These categories are ordered from the worst to the best. The outcome of a MCDA sorting method is an assignment of the alternatives among the different classes.

3.3   Multiple criteria decision aiding techniques

Three main types of approaches to multiple criteria decision aid are distinguished in the literature: functional, relational and logic models. Each one represents the user preferences in a different way.

Logic model or decision rules. Logic conditions are given by rules that permit the classification of the alternatives in some predetermined categories, taking into account the performance of the alternative with respect to the different criteria. In this case, the rules are the core of the preference model [28].

Functional models. A values system is used to associate marginal preferences upon each criterion to each of the reference indicators that describe the alternatives. These value functions permit the rating of each alternative according to its performance [28], [44]. All the value functions must generate a utility score in the same reference scale because later these partial utilities are aggregated to obtain the overall utility score of each individual alternative.

Relational models or outranking relations. Preferences are expressed as binary relations between the alternatives. Each criterion defines a partial preference structure on the set of alternatives, with three types of relations: indifference, weak preference and strong preference. From the analysis of this partial preference structure, an overall preference structure among the alternatives can be derived. The exploitation of the overall preference relations will lead to the solution of the decision problem. In this case, all alternatives are treated together.

3.3.1   Decision rules

The basic assumption of the decision rule approach is that it is usually easier for decision makers to give preferential information in terms of decision examples than to define criteria models. In some cases, we can

3.3. Multiple criteria decision aiding techniques 49

collect a database of past decision cases from which one can seek simple rules that justify the decisions. Some automatic procedures have been defined to extract such rules automatically. The most well-known is the *Dominance-based Rough Set Approach* (DRSA) [28].

In this framework the induced decision rules are made up of two parts of the form "If *condition*, then *decision*":

- A 'conditional' part that makes it possible to compare the values of one alternative with thoses of a reference once (i.e. conditions) by using the progressive dominance relation.

- A 'decisional' part that assigns a global preference category or that gives the preferred binary relation over the two alternatives based on a set of conditions stated in the premises.

In DRSA a fixed set of preference categories are associated to the decision variable. The decision makers must provide preferential information in the form of examples. Each example consists on a list of attribute-value pairs together with its corresponding preference category. Then, different types of rules can be constructed :

- *If* the alternative performs x in g_1 and y in g_2 and ... *then* it is comprehensively **at least** in category C_i

- *If* the alternative performs x in g_1 and y in g_2 and ...*then* it is comprehensively **at most** in category C_j

- *If* the alternative a is weakly preferred to b in g_1 and a is strongly preferred to b in g_2 and ...*then* a is **at least as good as** b

New sorting problems can be easily solved by applying the rules induced from the examples to the data of new alternatives.

Advantages and disadvantages of decision rules

One of the main advantages of this approach is that the obtained model is easy to understand for the DM because people is used to work with logic rules expressed in a linguistic way. Another advantage is the possibility of handling inconsistencies in the preferential information provided in the

examples. Inconsistencies may be solved by defining kernel criteria and rules with different levels of generality.

The main drawback of this method is the need of a set of examples of decision problems that have been solved previously under the same conditions. There are unique decision problems in which this previous experience is not available.

3.3.2   Multi-Attribute Utility Theory

The theory of *multi-attribute utility* [45] known as MAUT, is based on the idea of associating a utility rating, generally a real-valued number $u_j(a)$, to each alternative a on criterion j, representing the degree of "satisfaction" of a on criterion g_j, according to the DM's expectations. The theory of utility includes functions that represent the preferences of decision makers on alternatives after an evaluation. In that way, a utility function $u_j : A \rightarrow S$ is defined for each criterion to transform the original scores of the alternative on a certain variable into satisfaction scores.

Once a real-valued function u_j is defined for all criteria, for each alternative a procedure of aggregation of the corresponding partial utility ratings into a global utility score is done. Aggregation is represented with a function $H : \mathcal{S}^m \rightarrow \mathcal{S}$. Several aggregation operators have been proposed in the literature, requiring the following mathematical properties [15]:

1. Idempotency: $H(r, r, ..., r) = r, \forall r$

2. Monotonicity: $r_j' > r_j \Rightarrow H(r_1, ...r_j', ...r_m) \geq H(r_1, ...r_j, ..., r_m)$

3. Commutativity: $H(r_1, r_2, r_3) = H(r_{\sigma(1)}, r_{\sigma(2)}, r_{\sigma(3)})$ being σ a permutation of the values.

4. Boundaries: $\overset{m}{\underset{j=1}{\wedge}} r_j \leq H(r_1, r_2, ..., r_m) \leq \overset{m}{\underset{j=1}{\vee}} r_j$

5. Associativity: $H(r_i, r_j, r_k) = H(H(r_i, r_j), r_k), \forall i, j, k$

6. Decomposability: $H(r_1, r_2) = r' \Rightarrow H(r_1, r_2, ..., r_m) = H(r', r', r_3, ..., r_m)$

3.3. Multiple criteria decision aiding techniques   51

Classic aggregation operators

For a comprehensive review of numerical aggregation operators, the reader is referred to [15], [89], [90]. Let us just outline the most common ones:

1. Quasi-arithmetic means: Represents the family of means which include simple arithmetic mean, geometric mean and harmonic mean, among others. It is defined as follows:

$$H(r_1, r_2, ..., r_m) = f^{-1}\left[\frac{1}{m}\sum_{j=1}^{m} f(r_j)\right]$$

It can be extended to apply a set of weights w_j as follows:

$$H(r_1, r_2, ..., r_m) = f^{-1}\left[\sum_{j=1}^{m} w_j f(r_j)\right]$$

2. Median: Applying the concept of median in statistics, the real-valued function r_j of the alternatives is not taken into account but only their ordering. The median is the middle value of this ordered list.

$$H(r_1, r_2, ..., r_m) = \begin{cases} r_{\left(\frac{n+1}{2}\right)} & \text{if } m \text{ is odd,} \\ \frac{1}{2}\left(r_{\left(\frac{n}{2}\right)} + r_{\left(\frac{n}{2}+1\right)}\right) & \text{if } m \text{ is even} \end{cases}$$

3. Ordered weighted averaging operators (OWA): It provides a parameterized class of mean-type aggregation operators. It establishes a trade-off between conjunctive and disjunctive models of aggregation by means of a value-based weighting vector ω. OWA generalizes other mean operators such as the max, arithmetic average median and min. It was introduced by Yager 1998 [97]. It is formally defined as follows:

$$H(r_1, r_2, ..., r_m) = \sum_{j=1}^{m} \omega_j b_j$$

where b_j is the $j-th$ largest of the rating in $r_1, ..., r_m$ and $\sum_{j=1}^{m} \omega_j = 1$.

It is worth noting that the WOWA operator used in the previous chapter is also an aggregation operator satisfying the properties indicated before. In this case, it is using at the same time criteria weights w and value-based weights ω.

The global utility obtained for each alternative on A allows their comparison and the construction of a ranking of alternatives on A from the best to the worst in a complete, transitive pre-order. Following the trade-off nature of this approach, the global utility is always comparable between alternatives, so alternatives cannot be incomparable with this approach.

Advantages and disadvantages MAUT

The theory of multi-attribute utility associates a numerical utility score to each alternative. This approach has the following advantages and disadvantages:

Advantages

- The performance of the criteria is trasformed into a common scale S (for example, monetary scale, satisfaction scale), which facilitates the comparison of alternatives.

- The alternatives obtain a final overall score that naturally leads to a total preorder which is easy to interpret.

Disadvantages

- The complexity of the axioms requires a cognitive effort on the side of the decision maker.

- Sometimes the utility function is difficult to define and interpret, especially when the attributes are of very different natures.

- Aggregation is usually compensatory, such that a poor performance on one attribute can be offset by a good performance on another.

3.3.3 Outranking methods

The outranking approach was developed in France by Bernard Roy and it has been widely applied in Europe [73], [64]. The principle of this

3.3. Multiple criteria decision aiding techniques 53

approach is to make a pairwise comparison of all alternatives on each attribute in order to build binary preference relations. The relations found on each individual criterion are then used to construct an overall binary outranking relation, which can be Boolean or fuzzy.

Roy defined an outranking relation as a binary relation aSb on the set of alternatives such that it is satisfied when there are enough arguments to declare that a **is at least as good as** b, while there is no essential reason to refute that statement. The arguments taken into account are the preferences of the decision-maker, the quality of the evaluations of the alternatives and the nature of the problem.

Outranking methods follow two phases: first, a construction phase where the outranking relation is built and, second, a relation exploitation phase to obtain a final recommendation for the decision maker [74]. The strength of the assertion a outranks b is given by a credibility or outranking degree $S(a, b)$. It is a score between 0 and 1, where the closer $S(a, b)$ is to 1, the stronger the assertion. This outranking degree $S(a, b)$ considers two perspectives : the concordance and the discordances of the statement that a outranks b. Thus, outranking relations are subject to two conditions that represent these ideas [67]:

- Condition of concordance: in each pair, a majority of criteria must emerge in favor of the best alternative.

- Condition of discordance: in each pair, in the minority of criteria, there must not be too much pressure in favor of a reverse upgrade; in other words, no criterion should veto the best alternative.

The DM is required to express his/her preference information on the attribute scale by means of two thresholds: indifference and preference. They are used to construct the criterion by which the performance of the alternatives will be evaluated in terms of concordance and discordance.

The main methods withing the outranking approach are ELECTRE and PROMETHEE. Both of them have different versions to solve choice, ranking and sorting problems. In this thesis, we will study in detail the ELECTRE method because it includes some aspects that are not present in PROMETHEE. The next chapter will be devoted to ELECTRE and the integration of semantic criteria in its procedures. Here we will just briefly outline the other method, PROMETHEE.

PROMETHEE method

The PROMETHEE (*Preference Ranking Organization MeTHod for Enrichment Evaluations*) method was first proposed in [13]. It builds a valued outranking relation based on a preference index $P_j(a, b) \in [0, 1]$ representing the degree of preference of a over b for each criterion on G. It is calculated from the difference between the performance of the alternatives, so that $P_j(a, b) = f(g_j(a) - g_j(b))$. The closer $P_j(a, b)$ is to 0, the greater the indifference between a and b is, while the closer it is to 1, the greater the preference of a over b is. Note that this preference index gives a valued "preference degree" between two alternatives.

This preference index can be defined in different ways.

1. Usual criterion: The indifference only applies when $g_j(a) = g_j(b)$. Otherwise, the DM is indicating a strict preference of the alternative with the best performance.

2. Quasi criterion: The criterion is associated to a threshold q. If the difference between $g_j(a)$ and $g_j(b)$ does not exceed this threshold, then a and b are indifferent. Otherwise, the alternative with the best performance is strictly preferred.

3. Criterion with linear preference: The function is associated to a threshold p. If the difference between $g_j(a)$ and $g_j(b)$ is lower than p, the DM is indicating a progressive preference of the best performance. Otherwise, it is strictly preferred.

4. Level criterion: In this function, the DM has to set the two thresholds q and p. If the difference between $g_j(a)$ and $g_j(b)$ does not exceed q the alternatives are indifferent. If it is between q and p there is a weak preference (0.5). After this value, it becomes a strict preference of the alternative with the best performance.

5. Criterion with linear preference and indifference area: In this function, a and b are considered indifferent as long as $g_j(a) - g_j(b)$ does not exceed q and the preference increases linearly from q until p. After p, the strict preference applies.

6. Gaussian criterion: This function (ρ) is made easily using the normal distribution in statistics.

3.3. Multiple criteria decision aiding techniques 55

Assuming that, for all pairs of alternatives $(a, b) \in A$, the preference indices $P_j(a, b)$ have been calculated, the overall preference $\Pi(a, b)$ is calculated taking into account a weight w_j of each criterion j. This preference $\Pi(a, b)$ represents the weighted average of the partial preference functions $P_j(a, b)$. It is calculated as follows:

$$\Pi(a, b) = \frac{\sum_{j=1}^{m} w_j P_j(a, b)}{\sum_{j=1}^{m} w_j} \tag{3.1}$$

These values are the scores associated to the outranking relation aSb, therefore they are used in the next stage to solve the decision problem. There is no possibility of veto in PROMETHEE, so the minority condition is not considered in this case.

The preference indices Π for all pairs in A are represented as a valued graph. This graph is represented by modeling two arcs between alternatives a and b, representing $\Pi(a, b)$ and $\Pi(b, a)$ respectively. For a certain alternative we can define two concepts based on these arcs: entering flow and leaving flow. These flows represent the origin and destination, so that for instance, the arc represented by $\Pi(a, b)$ indicates that an arrow is leaving from a and entering in b.

The leaving flow of node a is the sum of the arcs leaving a, providing a measure of the outranking character of a. It is calculated as follows:

$$\eta^+(a) = \sum_{b \in A} \Pi(a, b) \tag{3.2}$$

The entering flow of a measures the outranked character of a. It is calculated as follows:

$$\eta^-(a) = \sum_{b \in A} \Pi(b, a) \tag{3.3}$$

Using these positive and negative flows for all alternatives in A, different exploitation procedures of this graph can be applied to provide the best solution depending on the kind of problem that the DM is facing (i.e., ranking, sorting or choice). The two most well-known PROMETHEE methods are PROMETHEE I and PROMETHEE II, which are applied to ranking problems. For example, in the case of PROMETHEE I, the partial pre-order is obtained from the entering and leaving flows as follows:

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

56   Chapter 3.   Multiple Criteria Decision Aiding

- aPb: if $\eta^+(a) > \eta^+(b)$ and $\eta^-(a) < \eta^-(b)$, or
  $\eta^+(a) > \eta^+(b)$ and $\eta^-(a) = \eta^-(b)$, or
  $\eta^+(a) = \eta^+(b)$ and $\eta^-(a) < \eta^-(b)$;

- aIb: if $\eta^+(a) = \eta^+(b)$ and $\eta^-(a) = \eta^-(b)$;

- aRb: otherwise.

Advantages and disadvantages of the outranking approach

Advantages

- Heterogeneous scales: the criteria can be evaluated on different types of scales without the use of any normalisation procedure.

- Non-compensatory nature: unlike utility theory, which allows trade-offs between criteria, outranking methods do not allow the compensation between the performance criteria; in other words, performance degradation on some criteria can not be offset by performance improvements on other criteria.

Disadvantages

- Outranking methods are not very suitable for problems involving a large number of alternatives or attributes because they are computationally expensive [20].

- The definition of the thresholds is sometimes a hard task for the decision maker.

3.4   Summary

In this chapter we have presented the basic concepts and models of multi-criteria decision aiding methods. As this thesis is focused on the outranking methods, we have introduced the basic concepts of this approach together with the PROMETHEE method. It can be seen that PROMETHEE does not consider the opinion of minorities, which is a limitation that may have important consequences in certain types of problems, because a unique strong reason opposing to the majority of criteria may be worth taking into

account. This is especially critical when criteria representing very different points of view must be merged together. For this reason, the ELECTRE method is the one studied in this dissertation.

The introduction of semantic criteria in each of the three MCDA methodologies has to be studied independently, as each one follows a very different approach. In decision rules semantic criteria have not yet been used, as far as we know. It would be possible to study how to manage the list of tags and its associated scores by means of a rough set approach. Regarding the second model, the multi-attribute theory based on utility functions, the tag interest scores can be understood as utility scores, therefore an aggregation of all the TIS of one criterion g_j of a certain alternative a could be done to obtain its corresponding partial utility $r_{(a)}$. This has been already proposed in [11]. The main disadvantage of this approach is the loss of information during the aggregation of all the TIS into a unique value. The next chapter describes the new ELECTRE-SEM method, which allows the use of semantic criteria in the definition of alternatives.

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

4

ELECTRE-SEM: an outranking method with semantic criteria

As will be described in ths chapter, an important advantage of ELECTRE is that it can work directly with the original attribute scales, without requiring their transformation into abstract ones with an arbitrary range. However, one of the main shortcomings of ELECTRE is that it assumes that criteria are numerical or ordinal, but not semantic. In this dissertation the classic ELECTRE methodology has been enhanced so that it can handle semantic criteria. The preferences of the user are stored in an ontology-based user profile, as described in chapter 2. There is a personalized ontology for each user, which contains the degree of preference of the user with respect to the most specific concepts of the ontology (TIS values). In this chapter, we will explain how this preferential information may be exploited to compare and rank a set of alternatives.

4.1   The classic ELECTRE method

The ELECTRE (*ELimination Et Choix Traduisant la REalité*) methodology was designed in France in 1965 by Bernard Roy [73] in order to tackle new decision problems. Different methods have been defined from the original one, each one tackling a different kind of problem or extending the initial definitions with more general ones, adding more flexibility. At the moment, the most well-known methods of this family are ELECTRE I and

Is (for choice problems), ELECTRE II, III and IV (for ranking problems) and ELECTRE-Tri-B and Tri-C (for sorting problems) [28].

The main goal of the ELECTRE methods is to establish one of these four basic situations of preference for each pair of alternatives (a,b):

- Indifference (aIb): it corresponds to a situation where there are clear and positive reasons that justify an equivalence between the two actions.

- Strict preference (aPb): it corresponds to a situation where there are clear and positive reasons in favour of one of the two alternatives.

- Weak preference (aQb): it corresponds to a situation where there are clear and positive reasons that invalidate strict preference in favor of one of the two alternatives, but they are insufficient to deduce either the strict preference in favour of the other alternative or indifference between both alternatives, thereby not allowing either of the two preceding situations to be distinguished as appropriate.

- Incomparability (aRb): it corresponds to an absence of clear and positive reasons that would justify any of the three preceding relations.

In that way, the ELECTRE outranking method builds a reflexive, non-transitive preference relation, S, between potential alternatives. Given two alternatives a and b, aSb means "a is at least as good as b". The outranking relation is true if and only if there are enough criteria that support this statement and no criterion refutes it. Therefore, criteria are seen as voters in favor or against the claim "a is at least as good as b". Criteria in *concordance* are the ones that support it, while criteria in *discordance* are the ones refusing it. This mechanism is inspired in electoral procedures based on voting techniques from the social choice field.

It is possible to relax the model by allowing a valued outranking relation, instead of a simple Boolean one. In that case, the concordance and discordance statements are fuzzy and they are calculated using some numerical indices. This situation is used to model the decision maker's uncertainty and imprecision associated to the pairwise comparison of the alternatives, which is handled with the addition of the following discrimination thresholds:

- Indifference threshold $q_j(a)$: given two alternatives a and b, it is the maximum difference of the scores on criterion g_j below which the decision maker is indifferent between both options.

- Preference threshold $p_j(a)$: given two alternatives a and b, it is the minimum performance difference of the scores on criterion g_j which implies a clear strict preference in favour of one alternative over another.

Depending on the values of the discrimination thresholds, we can find three situations with different level of generality: true-criteria, quasi-criteria and pseudo-criteria.

- True-criteria: This criterion model applies to a criterion g_j when $q_j(a), p_j(a) = 0$. Thus, indifference only occurs when $g_j(a) = g_j(b)$.

- Quasi-criteria: This criterion model considers indifference between small differences, such that $q_j(a) > 0$ and $q_j(a) = p_j(a)$.

- Pseudo-criteria: The most recent ELECTRE methods model criteria as pseudo-criteria [71] for handling the imprecision and uncertainty inherent to complex human evaluation processes. Consequently, the outranking relation can be interpreted as a fuzzy relation, such that $q_j(a), p_j(a) > 0$ and $q_j(a) < p_j(a)$.

4.2 The ELECTRE algorithm

In this dissertation we have focused our attention on the ELECTRE-III ranking methodology, which follows two main steps (Figure 4.1):
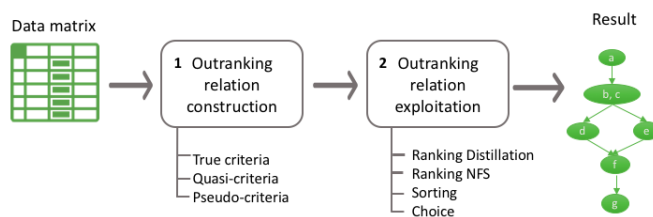
Figure 4.1: Outranking relation construction and exploitation

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

Chapter 4.   ELECTRE-SEM: an outranking method with
62   semantic criteria

4.2.1   Step 1 - Construction of the outranking relation

The outranking relation S is built for each pair of alternatives (a, b) by comparing their performance on the set of criteria G. The alternative a outranks the alternative b (aSb) if, taking into account the decision makers preferences, a is at least as good as b and there is not any strong argument against this claim. As said before, two indices are applied to evaluate this relation on each criterion: concordance and discordance. From the values of these indices it is possible to compute the overall degree of credibility of (aSb). These measures are defined in the following paragraphs.

Partial Concordance Index

For each criterion $g_j \in G$, the *partial concordance* $c_j(a, b)$ is calculated using the indifference and preference thresholds as follows:

$$
c_j(a, b) = \begin{cases} 1 \text{ if } g_j(a) \geq g_j(b) - q_j(b) \\[2mm] 0 \text{ if } g_j(a) \leq g_j(b) - p_j(b) \\[4mm] \frac{g_j(a) - g_j(b) + p_j(b)}{p_j(b) - q_j(b)} \text{ otherwise.} \end{cases} \tag{4.1}
$$

Overall Concordance Index

Once the partial concordances have been measured, an overall concordance index is computed for each pair of alternatives (a, b) as follows:

$$
c(a, b) = \frac{1}{W} \sum_{j=1}^{m} w_j c_j(a, b) \tag{4.2}
$$

In this expression w_j is the weigth of criterion g_j and W is the addition of the weigths of all the criteria.

Partial Discordance Index

ELECTRE-III also includes the *veto* rule, which is the right of giving essential reasons for rejecting the outranking relation. This is considered the respect to minorities. It is introduced as another threshold associated to the performance values of a criterion:

- Veto threshold $v_j(a)$: given two alternatives a and b, a discordant difference larger than the veto in favour of b with respect to a in criterion g_j will require the negation of the outranking relation aSb

(thus, if there is a criterion in which b is much better than a, it will not be possible to claim that a is at least as good as b).

The partial discordance index is defined as:

$$d_j(a,b) = \begin{cases} 1 \text{ if } g_j(a) \leq g_j(b) - v_j(a) \\ 0 \text{ if } g_j(a) \geq g_j(b) - p_j(a) \\ \\ \frac{g_j(b) - g_j(a) + p_j(a)}{v_j(a) - p_j(a)} \text{ otherwise.} \end{cases} \tag{4.3}$$

Outranking credibility value

Finally, the degree of credibility of the outranking relation aSb, $\rho(a,b)$, is calculated using the global concordance and the partial discordance indices of the set $J(a,b)$ of criteria for which the discordance is larger than the overall concordance.

$$\rho(a,b) = \begin{cases} c(a,b) & \text{if } \forall_j d_j(a,b) \leq c(a,b), \\ c(a,b) \prod_{j \in J_{(a,b)}} \frac{1 - d_j(a,b)}{1 - c(a,b)} & \text{otherwise} \end{cases} \tag{4.4}$$

4.2.2 Step2 - Exploitation of the outranking relation

The outranking relation is exploited in the second stage. The exploitation procedure depends on the type of problem: choice, ranking or sorting. In each case, different procedures are needed because the answer to be given to the decision maker is very diverse. In a choice problem, we must find a subset of the best alternatives. This subset must contain only incomparable alternatives in terms of the S relation, and the rest of alternatives must be outranked by the chosen ones. In ranking problems, the DM wants to know the position of each alternative with respect to the others, in a total or partial ranking. Finally, in sorting problems, a predefined set of ordered categories is defined and the alternatives must be assigned to the corresponding category depending on the S relations.

In this work we have focused on the ranking case. The following two exploitation procedures are used in ELECTRE to build a ranking.

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

Chapter 4.   ELECTRE-SEM: an outranking method with
semantic criteria

64

Ranking method 1 - Net Flow Score (NFS)

The matrix that contains the credibility values for each pair of alternatives may be interpreted as the adjacency matrix of a labelled directed graph, in which the nodes are the criteria and the label of the edge between two criteria a and b is the value $\rho(a, b)$. The *Net Flow Score* (NFS) procedure (Szelag et al., 2014) analyzes this graph to calculate two evidences: strength and weakness. The *strength* of an alternative a is defined as the sum of the credibility values of the edges that leave from the node a. The *weakness* of an alternative a is defined as the sum of the credibility values of the edges that reach the node a. The NFS of an alternative a is the difference between its strength and its weakness. This value allows ranking the alternatives in a descending partial order.

$$NFS^{sval}(a) = \sum_{b \epsilon A}[S(a, b) - S(b, a)] \tag{4.5}$$

Ranking method 2 - Distillation

In the so-called ELECTRE-III version a *distillation* procedure that exploits the outranking relation to build a partial pre-order among the alternatives in A was defined. It is an iterative process that selects at each step a subset of alternatives, taking into account the credibility values of the outranking relation. This procedure yields two complete pre-orders (descending and ascending distillation chains, O_d and O_a), which are intersected to generate the final partial pre-order [40]. The descending distillation procedure is the following:

- Starting from the complete set of alternatives, the ones with a highest qualification are extracted to form a first group ($Distillate_1$). This set is found by using a cut-off level of credibility, λ. The alternatives in this group must be indifferent since it is not possible to establish a preference between them. This group is placed at the top of the ranking.

- From the remaining set of alternatives, the best ones are again extracted to obtain a second group ($Distillate_2$). In this step, and

the successive ones, λ is progressively reduced in order to make the condition of preference weaker.

- This procedure is repeated until all the alternatives belong to a distillate group.

Analogously, the ascending distillation procedure follows the same steps, but starting with the extraction of the worst subset (which is placed at the bottom of the ranking) and continuing upwards.

The intersection of these two pre-orders O_d and O_a gives a partial pre-order O. The rules of intersecton are based on finding a global relation between a pair of alternatives, observing their relations in these two pre-orders. If the relation is equal, it is established as the global one. In case of them being indifferent in one of the partial pre-orders but not in the other, the non-indifferent relation is taken as the global one. When the preference relation is opposite in the two pre-orders, then the alternatives are considered to be globally incomparable.

4.3   Managing semantic data with ELECTRE-SEM

In this section we propose a novel way to construct the outranking relation when semantic criteria are considered, called ELECTRE-SEM. In order to manage a different type of criteria we only need to change the first step of the ELECTRE method, because the exploitation procedure only takes into account the credibility values. In particular, it is necessary to define how to calculate the partial concordance index and partial discordance index in the case of a semantic criterion. Moroever, we will assume that these attributes are multi-valued, so each alternative will have an associated list of values for each semantic criterion. This aspect must also be taken into account in the definition of the indices.

The values that a semantic criterion may take will be represented in a domain ontology. Every linguistic value (tag) that appears on the semantic criteria is an elementary concept (i.e. a leaf) in the corresponding ontology. As described in chapter 2, the ontology will store a numerical preference score on each of these concepts according to the preferences of each individual user, called Tag Interest Scores (TIS). In the ELECTRE-

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

Chapter 4.   ELECTRE-SEM: an outranking method with
66   semantic criteria

SEM method, this information will be exploited to compare and rank the set of alternatives.

In short, the decision procedure consists of the following steps (see Figure 4.2):

1. The decision maker constructs his/her ontology-based subjective semantic user profile.

2. The data matrix is collected with the objective information corresponding to each alternative and criterion.

3. The parameters of the method (discrimination thresholds and criteria weights) are set up by the decision maker.

4. Concordance and discordance indices are calculated and a final ranking procedure is applied to obtain a partial pre-order or a total ranking, which is presented to the decision maker.
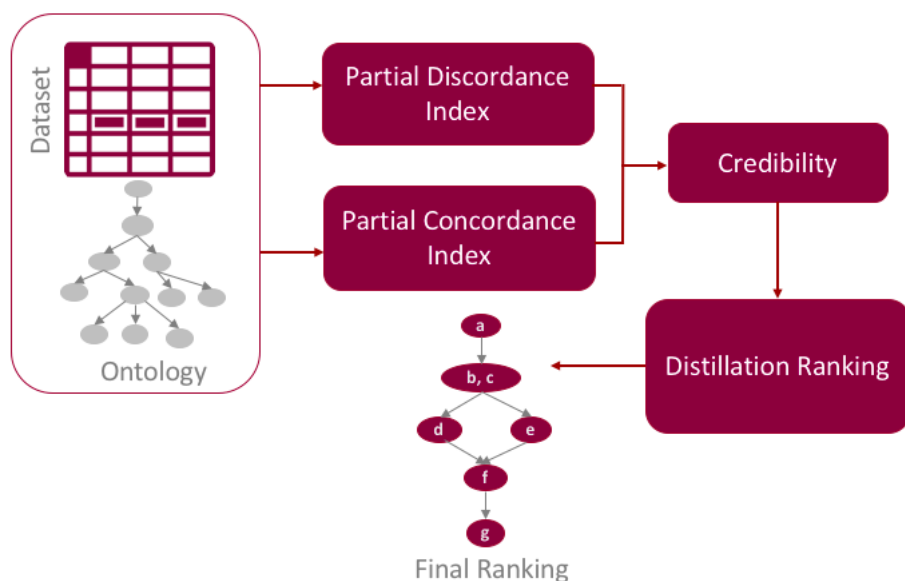
Figure 4.2: ELECTRE-SEM steps and data

4.3. Managing semantic data with ELECTRE-SEM 67

Each semantic criterion is defined as a pseudo-criterion, with two discriminant thresholds (preference and indifference) as well as the veto threshold. This procedure follows the same principles than the classic ELECTRE method, but concordance and discordance indices are fuzzy functions defined in terms of the pairwise comparison of the Tag Interest Scores. First, we define how to measure the strength of the assertion aSb in terms of one semantic variable:

Definition 6. *Semantic Win Rate* $SWR_j(a, b)$

It is a numerical value in $[0..1]$ *that indicates the degree of performance of the alternative a with respect to the alternative b on the semantic criterion* g_j. *It is based on the two sets of tags* $g_j(a) = \{t_{1,a}, t_{2,a}, t_{3,a}, \ldots, t_{|g_{j(a)|,a}}\}$ *and* $g_j(b) = \{t_{1,b}, t_{2,b}, t_{3,b}, \ldots, t_{|g_{j(b)|,b}}\}$ *(the values taken by the alternatives in the criterion) and it is calculated as follows:*

$$SWR_j(a, b) = \frac{\sum_{t_{i,a}} \sum_{t_{k,b}} f(t_{i,a}, t_{k,b})}{|g_j(a)| \cdot |g_j(b)|} \tag{4.6}$$

, where

$$f(x, y) = \begin{cases} 1 \text{ if } TIS(x) \geq TIS(y) - q_j \\ 0 \text{ if } TIS(x) < TIS(y) - q_j \end{cases} \tag{4.7}$$

Thus, $SWR_j(a, b)$ is the percentage of pairwise comparisons between the tags of a and b for the semantic criterion g_j for which the user has a higher (or equal) preference for the $a-tag$ than for the $b-tag$. We introduce here the possibility of using an indifference threshold q_j similar to the one in standard ELECTRE, in order to define an interval of indistinguishability regarding the TIS range of values. In that way, if two scores are similar enough, they can be considered equally preferred by the decision maker. It should be noted that in some problems the TIS value may represent the assessment of the risk associated to each tag and alternative, instead of a positive preference score. In these cases it should be minimized in Equation 4.7.

Example. Let us consider two lists of tags describing a touristic activity, with their associated TIS value (to maximize):

- a: (picnic 0.3, beach 0.6, swimming 0.2)

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

Chapter 4. ELECTRE-SEM: an outranking method with semantic criteria

68

- b: (shopping 0.2, history 0.9, roman 0.4)

If we take $q_j = 0$, $SWR_j(a, b) = 4/9$ and $SWR_j(b, a) = 6/9$, so option b is preferable to a, because the tourist is more interested in historic places and roman culture than in swimming and having lunch in the beach.

Let us now introduce some indifference on the risk assessment value, with $q_j = 0.1$. Now, $SWR_j(a, b) = 5/9$ and $SWR_j(b, a) = 7/9$. This means that scores 0.2 and 0.3 are considered to be in the same level of preference (for picnic and shopping). The same situation arises in the comparison of 0.4 and 0.3 (for picnic and roman). Thus, the semantic win rate (SWR) changes, but option b is still better than a, because b has a tag that is much more preferred than the ones of a.

Using the Semantic Win Rate value, the partial concordance and discordance indices are defined as follows:

Definition 7. *Partial concordance and discordance indices for semantic criteria (see Figure 4.3).*

$$c_j(a, b) = \begin{cases} 1 \text{ if } SWR_j(a, b) \geq \mu_j \\ 0 \text{ if } SWR_j(a, b) \leq p_j \\ \\ \frac{SWR_j(a,b) - p_j}{\mu_j - p_j} \text{ otherwise.} \end{cases} \quad (4.8)$$

$$d_j(a, b) = \begin{cases} 1 \text{ if } SWR_j(a, b) \leq v_j \\ 0 \text{ if } SWR_j(a, b) \geq p_j \\ \\ \frac{p_j - SWR_j(a,b)}{p_j - v_j} \text{ otherwise.} \end{cases} \quad (4.9)$$

As $SWR_j(a, b)$ is a percentage that represents the comparison of the performance of a over b, the thresholds are not parameterized and they have the following meaning:

- μ_j is a strong threshold of the strength of $SWR_j(a, b)$ to consider maximum concordance with aSb.

4.3. Managing semantic data with ELECTRE-SEM   69
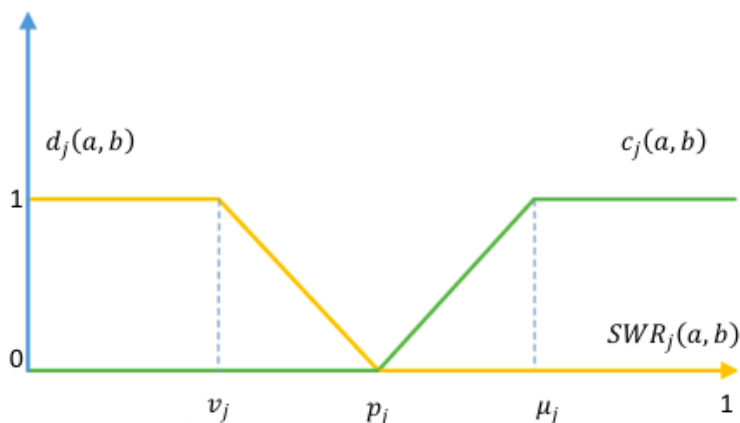
Figure 4.3: Fuzzy indices for concordance and discordance in semantic criteria

- p_j is a weak threshold of the strength of $SWR_j(a, b)$ where the user may still have some preference of a with regards to b, thus still supporting the relation aSb to a certain degree.

- v_j is the veto threshold, which is a value threshold below which $SWR_j(a, b)$ is low enough to imply the full discordance with the outranking relation.

In this case, the role of the thresholds is analogous to the one of the numerical case, p_j being the threshold that indicates if the value of the $SWR_j(a, b)$ is in favour of or against aSb, whereas μ_j and v_j are used to determine the value of the concordance or discordance vote for a certain criterion. Notice that the following condition must hold: $v_j \leq p_j \leq \mu_j$.

Example. Let us consider the same example shown above, in which $SWR_j(a, b) = 4/9$ and $SWR_j(b, a) = 6/9$. Let us consider two scenarios:

- First case: $\mu(j) = 5/9$, $p_j = 3/9$ and $v_j = 2/9$.

  In this situation, an option o_1 is fully preferred to another option o_2 if $SWR_j(o_1, o_2)$ exceeds $5/9$, and partially preferred if it is between $3/9$ and $5/9$. We totally disagree with the assertion that o_1 is better than o_2 if $SWR_j(o_1, o_2)$ is below $2/9$, and partially disagree if it is

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

70

Chapter 4.   ELECTRE-SEM: an outranking method with
semantic criteria

between 2/9 and 3/9. In this case, $c_j(a,b) = 0.5$ and $c_j(b,a) = 1$; thus, we fully support that b is better than a, and partially support that a is better than b. Regarding the discordance relationships, both $d_j(a,b)$ and $d_j(b,a)$ would be 0. Thus, there would not be any negative support on aSb or bSa.

- Second case: $\mu(j) = 7/9$, $p_j = 5/9$ and $v_j = 3/9$.

  In this scenario an option o_1 is fully preferred to another option o_2 if $SWR_j(o_1, o_2)$ exceeds 7/9, and partially preferred if it is between 5/9 and 7/9. We totally disagree with the assertion that o_1 is better than o_2 if $SWR_j(o_1, o_2)$ is below 3/9, and partially disagree if it is between 3/9 and 5/9. In this case, $c_j(a,b) = 0$ and $c_j(b,a) = 0.5$; thus, we only support (partially) that b is better than a. If we analyze the discordance relationships, $d_j(a,b) = 0.5$ and $d_j(b,a) = 0$. Thus, there would be some negative support on the assertion aSb.

4.4   Experiments

In this section we will analyze the behaviour of the new indices defined for semantic criteria, in particular in relation to the parameters of the model: weights and thresholds. We will use a case study related to the recommendation of touristic attractions in the city of Tarragona. Table 4.1 shows the 20 attractions that have been considered in this test. They are described using two criteria: a multi-valued semantic criterion that indicates the type of activity (*Touristic tags*) and a numerical criterion that indicates the price to pay (*Cost*). The user profile is defined in an ontology developed in the Scientific and Technological Park for Tourism and Leisure [60]. This ontology has 343 concepts, which are structured in 5 levels in a taxonomy.

The identifier of each alternative shows its stronger focus: C-Culture, E-Event, S-Sport and L-Leisure. We consider the case of a very sportive tourist, who has a mild interest in events and leisure activities but is not keen on cultural activities (except UrbanLandscape). It is assumed that all the preference scores of the ontology leaves have been calculated using the procedure described on chapter 2, from some basic initial information on the user's preferences.

4.4. Experiments 71

Table 4.1: List of alternatives for experiments with ELECTRE-SEM

ID	Touristic Tags	Cost
C1	CultureRoutes, Cathedral, Palace, Tower	30
C2	UrbanLandscape, CultureRoutes, HistoricBuilding, Tower, Baroque, Castle	30
C3	UniqueBuilding, Ruins, HumanHeritage, CultureRoutes, HistoricBuilding	30
C4	WineFairs, CultureRoutes, Ruins, Amphitheatre	15
C5	BookFairs, TraditionalCelebrations, HistoricBuilding	5
E1	TraditionalCelebrations, MusicFestivals, DanceFestivals, GastronomyFestivals	20
E2	WineFairs, MusicFestivals, ChampagneFestivals, BookFairs, DanceFestivals	20
E3	ChampagneFestivals, ArtsAndCraftsEvents, MusicFestivals, GastronomyFestivals	40
E4	BeachPicnic, DanceFestivals, BigGroupsAtmosphere, TapasCuisine, TraditionalCuisine	30
E5	TapasCuisine, ArtsAndCraftsEvents, BookFairs, TraditionalCelebrations, WineFairs	5
S1	Canoeing, Kayaking, BananaRafting, Windsurfing, WaterSkiing, Wakeboarding, ScubaDiving	80
S2	Snorkelling, Rappelling, ZipLine, BananaRafting, Kayaking	60
S3	HorseRiding, Car4x4, PaintBall, ShoppingArea	30
S4	SafariPark, HorseRiding, Car4x4, PaintBall	40
S5	Paragliding, ClimbingWall, Rappelling	10
L1	BeachPicnic, FamilyBeaches, Pizzeria, SafariPark	40
L2	TapasCuisine, ShoppingCenter, SpaResorts, Vegetarian, LocalMarket	20
L3	WineRoutes, TapasCuisine, WineFestivals, BookFairs	10
L4	Car4x4, PaintBall, HorseRiding, Pizzeria, SafariPark	40
L5	Bars, Discos, ShoppingArea, BeachPicnic, TraditionalCuisine, WineRoutes, SpaResorts	20

Test 1. This test shows the influence of the discordance index in the construction of the outranking relation. ELECTRE-SEM has been executed with or without discordance (i.e. without veto power) in each criterion. The parameters are the following: Cost (num, min, q=0, p=10, v=20) and Touristic Tags (semantic, max, μ =0.7, p=0.6, v=0.3).
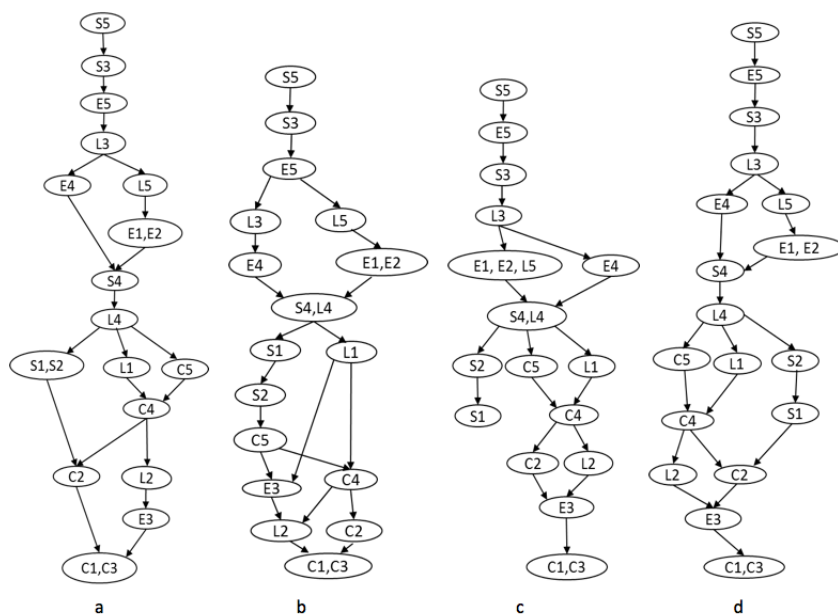
Figure 4.4: Results for Test 1

In this test both criteria have the same weight. The results are displayed in Figure 4.4 (a. partial pre-order with veto in both criteria, b. without veto in Cost, c. without veto in Touristic Tags, d. without any veto). In the first case we consider both the semantic information and the cost, so the best options are the cheapest sports (S5 and S3) and some cheap events and leisure activities. When the Cost veto is not considered the sports activities are promoted, so S1 and S2 are able to outrank more options (like C5, C4 and E3). In the third case, in which there is no veto for Touristic Tags, the price takes more importance and activities like S3 and S1 go down in the ranking. Finally, when no discordance is used (fourth

case), the result is very similar to the first one, because there are only two criteria. Notice that in all cases we identify some incomparability relations between activities that have good performance in one criterion but bad in the other (for instance between C5 -cheap but not in the interests of the user- and L1 -more interesting but much more expensive-). In the first positions, S3 (sport, 30) is better than E5 (event, 5) when there is no veto on cost, but the relation is reversed when there is no veto on the semantic criterion. These results show that the formulation of concordance and discordance indices for semantic data leads to plausible results when applied in the ELECTRE-SEM distillation procedure.

Test 2. This test studies the influence of the veto power (i.e. discordance) when there is a strong difference on the criteria weights (0.9 vs 0.1). The same previous 4 cases (with and without veto) have been studied, using the same parameters for the thresholds as in Test 1. The partial pre-orders are displayed in Figure 4.5.
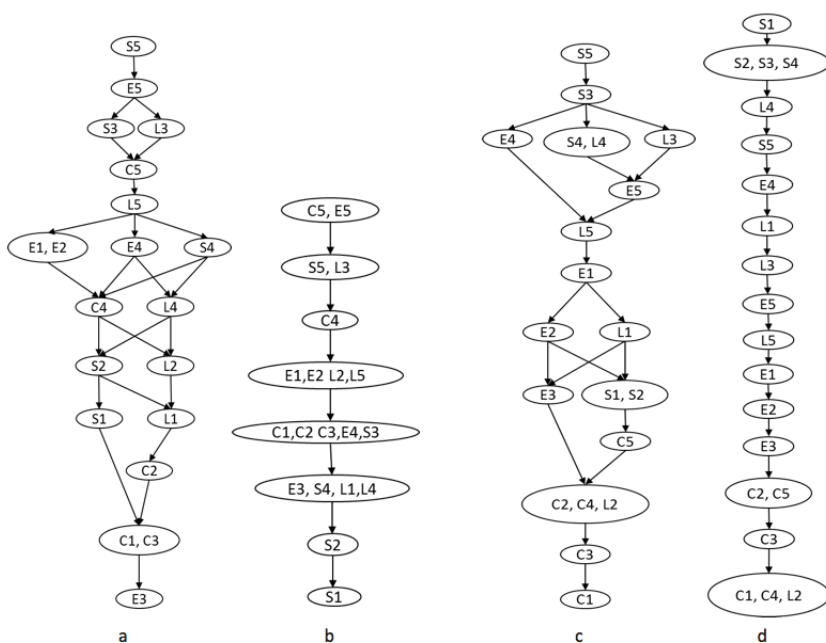
Figure 4.5: Results for Test 2

In the two figures on the left (a and b) $w_{cost}=$ 0.9 and $w_{tags}=0.1$, whereas in the two figures on the right (c and d) $w_{cost}=0.1$ and $w_{tags}=0.9$.

When the cost is the most important aspect, using or not its veto power (the discordance) leads to the same result, which is the first partial pre-order in Figure 4.5 (a). The right of veto of the semantic criterion is able to place S5 (10 euros) at the top, S3 (30 euros) in the third position and most cultural activities in the lowest ones, despite the extremely high importance of cost. However, when there is no veto in the semantic criterion (figure b) the ranking is mainly based on the cost and touristic tags are almost neglected, due to its low weight: C5 and E5 (5 euros) are the first options, S5 and L3 (10 euros) the second ones, etc. This test shows that the formulation of semantic discordance and concordance proposed in this work has the expected effect in the construction of the partial pre-orders. Similarly, when more importance is given to the semantic data, the ranking with cost veto (figure c) places S5 as the best alternative, S3 is the second one and (S4, E4, L3, L4) appear in the third place, because they are quite cheap and fit with the user's preferences. However, when there is no veto on cost (figure d) the ranking depends mainly on the scores of the touristic tags, which are based on the Semantic Win Rate, because the cost has an extremely low weight. Therefore, we can see that the veto power of the less relevant criterion is able to influence the result, both in the case on numerical data (classic procedure) and semantic values (new proposal).

4.5 Summary

Semantic information is nowadays frequent in many datasets and requires new analysis methods. In this chapter an extension of the ELECTRE multi-criteria decision-making method has been proposed. In particular, the procedure for constructing a fuzzy outranking relation is modified with the definition of new concordance and discordance indices. Those indices are calculated with a fuzzy function that depends on three thresholds (two for concordance and one for discordance) similar to those of the classic ELECTRE. However, they are based on the Semantic Win Rate, which is a new measure that permits to compare the lists of tags of a pair of alternatives taking into account the users preferences on the tags. In the next chapter of this dissertation we will describe how ELECRE-SEM has been implemented and how it has been applied to two decision problems in the Tourism and Environmental domains.

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

5

Tools and applications

In the previous chapter we have described the new framework ELECTRE-SEM, which allows the use of semantic criteria in the ELECTRE MCDA methodology. The management of semantic criteria has been included in a previous software (ELECTRE-H) developed by the ITAKA research group with the collaboration of researchers from the Poznan University of Technology in Poland. In this chapter we explain briefly this new version of the software and we describe two specific applications in domains which are considered relevant at URV: Environment and Tourism. In the first case, our new methodology has been applied to analyze several electricity generation technologies, considering environmental and economic criteria. This work was carried out during my research stay at Oxford Brookes University. In the second case, the new system has been applied to enhance the performance of a recommender system of touristic activities in the Tarragona province (which had been previously developed between the ITAKA research group and the Scientific and Technological park of Tourism and Leisure).

5.1   Software tool with semantic criteria

The *ELECTRE-H Software Package v1.0* was created by Luis Del Vasto, Aida Valls and two researchers of the Poznan University of Technology in Poland (Roman Slowinski and Piotr Zielniewicz) as a tool for multi-criteria decision aiding based on the outranking method ELECTRE-H, which is

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

78   **Chapter 5.   Tools and applications**

an extension of the classic ELECTRE method that may handle hierarchies of criteria. The decision maker may decompose the problem into smaller sub-problems, following the natural human decision making process. In this thesis we have extended the capabilities of this software tool by adding the management of semantic criteria. We have deployed a new system called *ELECTRE-H Software Package v2.0*, which is available in the Innoget software platform (via Universitat Rovira i Virgili).

This software package solves two types of problems: ranking and sorting. In the case of ranking, two methods are offered:

- ELECTRE-III-H: It was designed for finding a ranking of a set of alternatives according to the preferences of the decision-maker. The ranking is obtained at all levels of the hierarchy. It uses the exploitation method known as *distillation*, which computes partial preorders at all non-elementary nodes of the hierarchy, providing binary preferences among the alternatives at each sub-problem and at the most general problem. It is an extension of the classical ranking ELECTRE-III method.

- ELECTRE-NFS-H: It also generates a ranking at all levels of the hierarchy. The main difference between this method and the previous one is that the ranking procedure is based in the Net Flow Score technique, resulting in a total order ranking at each intermediate node, as well as in the root one.

In the case of sorting, the software package includes ELECTRE-TRI-B-H, which was designed for ordinal classification or sorting of alternatives into predefined categories at all levels of the hierarchy. It is possible to define different categories at each sub-problem and at the general problem, depending on their nature. It is an extension of the classical ranking ELECTRE-TRI-B method.

All these hierarchical methods introduce new formulations of the concordance and discordance indices to calculate the credibility of the outranking relation aSb at non-elementary criteria [25], [26].

The main user interface for ranking problems, which are the ones studied in this thesis, is shown in Figure 5.1. It has two parts: on the left hand side there is an area in which the hierarchy of criteria is shown, and on the right there is a panel to configure the parameters of each method.

5.1. Software tool with semantic criteria 79
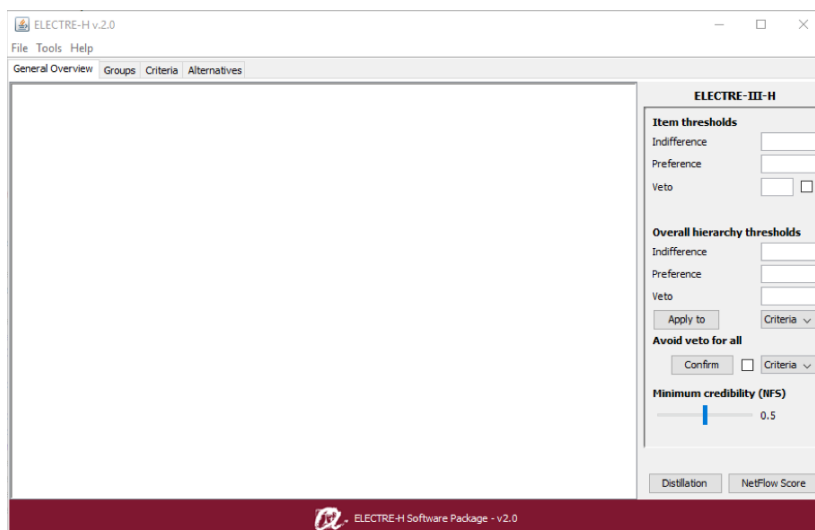
Figure 5.1: Main window of the ELECTRE-H software package v.2.0

Figure 5.2: Example of Excel input data file for ELECTRE-H v.2.0

The data is introduced in the software by means of an Excel file. All the data has to be in a single sheet, which has to include 3 sections for the following parameters: groups, criteria and evaluations of alternatives. Numerical, categoric and semantic types of data may be used (Figure 5.2). In the case of having semantic criteria, the input data is divided in 3 common main subsections, but we need to introduce some additional parameters. The software package includes the illustrative files "Example-SEM.xls", "ontologyTourismITAKA.owl" and "profileTourismITAKA.xls", which show an example that consists in ranking a set of touristic activities according to the decision maker preferences and two criteria, one semantic and the other numerical.

Figure 5.2 shows a part of the Excel file. In the lines in yellow there is the definition of a semantic (s) criterion, Tourist Activities, and a numeric (n) one, Cost. Each criteria has a weight and the values of the indifference, preference and veto thresholds. In addition to the thresholds, three more data are required:

- The value of the minimum SWR (Semantic Win Ratio), that must be a number in [0, 1].

- The path of the ontology associated to each semantic criterion. The ontology file must be written in OWL and have the extension .owl.

- The path of the profile Excel file. To facilitate the use of the software, the user profile consists in an Excel file with a table of tags and TIS, which stores the users preferences. The TIS (Tag Interest Score) must be a number in [0, 1]. The direction of the preference about the TIS can be given in the criterion description.

In the bottom part of Figure 5.2 we can see, in purple, the values of the criteria for some specific activities. Note that the semantic criterion Tourist Activities is multi-valued, so there is a list of tags for each activity.
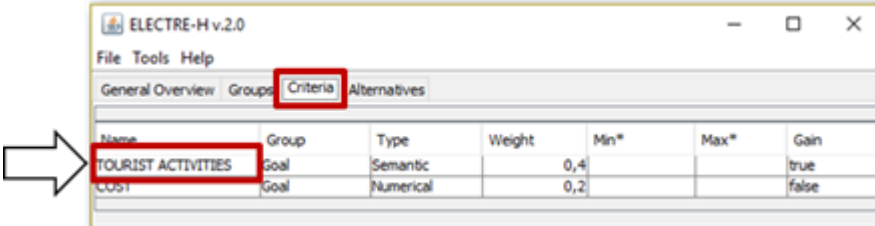
After loading the Excel file we can visualize the information content in the tabs Groups, Criteria and Alternatives. For semantic criteria, if the user clicks on the name of the semantic criterion (Figure 5.3), the system will display the ontology contents and it will offer some search options to visualize the user interest scores (i.e. the user profile), as illustrated

5.1. Software tool with semantic criteria 81

in Figure 5.4. It is also possible to change the size of the ontology or to obtain information about a concept. The user can also search for a certain concept and then only the related portion of the ontology is displayed. Moreover, it is also possible to filter the concepts by the TIS range.

Once the data has been loaded and all parameters have been configured, we must press the Solve button in the main window. A new window appears with the results. The Result Panel shows the ranking of all intermediate and root criteria in a descending order. In the Preorder panel, the Save button saves the image in PNG or JPG format (Figure 5.5).

Figure 5.3: List of criteria in ELECTRE-H v.2.0
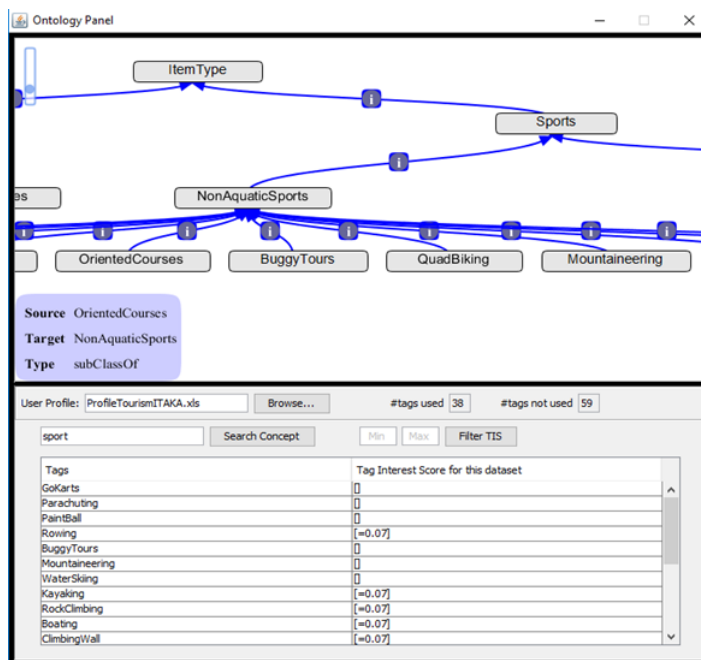
Figure 5.4: Ontology and user profile visualization in ELECTRE-H v.2.0
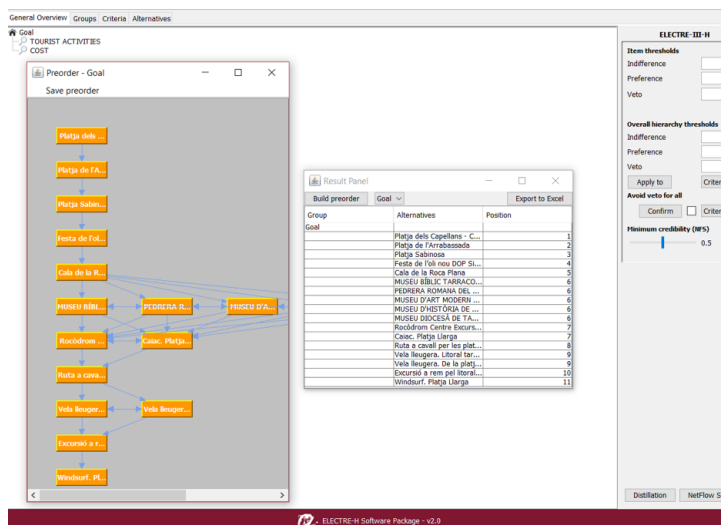
Figure 5.5: Visualization of results in ELECTRE-H v.2.0

5.2   Evaluation of energy generation technologies

5.2.1   Introduction

The selection of the most suitable electricity generation plant is a very controversial topic worldwide that requires the analysis of multiple factors and the consultation of many stakeholders. For instance, the decision of the British government to build new nuclear power stations has received a lot of criticism since its approval in 2008 and has taken many years of amendments and debates until the first one, Hinkley Point C, was licensed in September 2016. There are many conflicting facts that need to be considered in order to find the most appropriate technologies for electricity production on each site. On the one hand there is an increasing demand for energy, but on the other hand there is a great awareness for the need to protect the environment and reduce CO_2 emissions. In the literature, we can find several studies on sustainable energy production plants as a new means of generating energy while preserving the environment. The increase in concentration of atmospheric greenhouse gases (GHGs) due to the use of fossil fuels has led to the study of other energy supplies with lower GHG emissions as well as the use of renewable resources.

For a complete evaluation of all the possible technologies, several indicators must be collected and properly analysed by taking into account the concerns and aims of the stakeholders involved in each particular case (i.e. the criteria used to evaluate the different alternatives will be very different in Iran, Japan or the UK). Fortunately, nowadays it is possible to find the required information in big public databases. Detailed reports focused on each country are also widely available in public online data stores and can be included in the analysis. In this kind of complex decision problems, multiple and conflicting criteria must be taken into account, so MCDA methods can be applied.

This section will start with a review of research carried out on the application of multi-criteria decision support systems in the domain of electricity generation technologies. After that, we will explain how Electre-Sem has been used to assess several power generation technologies aimed to renovate the UK energy sector.

5.2.2   Related work

Most of the studies that evaluate electricity generation plants consider five categories of criteria: environmental, economic, technological, political, and social [93], [85] Technological considerations include efficiency, safety, reliability and resource availability. The main economic criteria are the costs of investment, operation, maintenance and fuel. Environmental criteria comprise VOC, resource depletion, noise, and the emission of NOX, CO2, SO2 and particulates. Social considerations include social acceptability, job creation and social benefits. Some papers give more importance to environmental and economic criteria and treat the others as complementary considerations [70], [78], [38].

Recent publications consider both renewable and non-renewable energies [3], [82], [39], [78], [72], [84], [21], [70], [83]. In other studies, only renewable technologies are taken into consideration [43], [63], [32], [35], [31], [85].

In all these studies, decision aid methods have been used to make an integrated analysis of the different energy generation technologies. Some approaches rely heavily on the knowledge and participation of a set of experts, such as the Delphi method [43]. In other papers, typical economic tools are used, for example the DEA method [72]. In the studies where MCDA methods are applied, the most common approach is based on utility theory, mainly using the Analytic Hierarchy Process (AHP) [21], [31]. Outranking methods have also been explored, such as PROMETHEE [35] and ELECTRE [63] but they are limited to numerical data, because they did not allow linguistic information until now.

The use of semantic information in decision support systems is an incipient research line. A usual and succesful way of introducing this kind of knowledge is using ontologies [41]. In [92] several ways of using the linguistic terms stored in the ontology for the description of objects are presented.

Another approach consists of using the ontology to represent the procedural and managerial information about a certain domain (i.e. tasks and their dependencies, requirements, resources, etc.). This latter approach is used for project management, strategic planning or to represent information and knowledge to facilitate system decision-making [47], [1] with renewable energy technologies.

Compendium [80] is a software tool based on the Issue Based Informa-

5.2. Evaluation of energy generation technologies 85

tion System (IBIS) [48]. Compendium allows information and ideas to be linked together through a visual interface. These concepts are expressed in the form of issues (question nodes), potential solutions (answer/position nodes) and arguments (pros and cons nodes). In OUTDO [39], an extension of Compendium that encapsulates a MCDA is used [4]. The modified Compendium system supports the decision-making process by integrating a qualitative representation of the argumentation and rationale behind the different alternatives with quantitative criteria that evaluate them. The amended Compendium system, OUTDO [39] was used to evaluate different electricity generation processes by considering diverse energy policies aimed to renovate the UK energy sector. The study focuses on nuclear power, coal with carbon capture and storage, and renewable energy generation. The case study presented in this section has been inspired by and is based on the study of UK power production plants developed by Hunt. The alternatives and some criteria have been taken from the case study developed in OUTDO. Semantic criteria have been added now using data available in public repositories, as they were not considered in the previous study. Moreover, we use an outranking-based approach, instead of a utility-based model as the one of OUTDO. Advantages of outranking methods, and ELECTRE in particular, have been previously recognized in the literature [29]. ELECTRE has 3 main advantages which are of interest in this study: different scales of measurement can be used without the need of a normalization pre-processing, compensation among criteria can be avoided with the veto power, and uncertainty in the performance comparison is managed by means of defining appropriate discrimination thresholds for each criterion. ELECTRE has been already successfully used in other environmental problems where the ELECTRE model may capture the complexity of the decision requirements in this domain [34].

In this case study there are two semantic criteria (waste by-products and pollution environmental damage) and the other three are numerical (energy source, economic cost of the electricity generation, and water usage). The aim is to identify the type of power generation plant that can best mitigate the effects it has on the environment. First we comment the technologies that will be evaluated and then we describe in more detail the considered criteria.

5.2.3   Technologies for energy generation

To preserve the environment and reduce CO2 emissions, governments are considering the incorporation of renewable energy, nuclear plants and new technologies to counteract climate change. The different types of energy sources currently available are classified into non-renewable and renewable.

Table 5.1: Energy generation technologies

Renewable	Sub-category
Solar Photovoltaic	Concentrated Photovoltaic (SP)
Wind	Offshore (WO)
Hydropower	Conventional-Aggregated In-Stream and Reservoir (HY)
Biopower (BIO)	
Non-renewable	
Natural Gas Combined Cycle (NGCC)	
Integrated Coal Gasification Combined (IGCC)	
Nuclear Power (NCL)	
Pulverized Coal (PC)	

1. Non-renewable sources [27], which have been used extensively until now, are nuclear fission, natural gas, and coal. One of the main disadvantages of coal power plants is the amount of pollution that the combustion of coal generates (NOX, CO2, and SO2). Nuclear power systems do not depend on renewable resources, but they reduce the consumption of fossil fuels (coal and oil) and have lower greenhouse gas emissions (CO2). However, nuclear waste is very radioactive, has a very long-life span and is difficult to safely dispose of.

2. Renewable energy that comes from sources such as wind, geothermal heat, sun, sea and organic waste have become popular in the last decades. Wind power systems do not produce harmful emissions; moreover, they cause minor disruption to the environment and they do not depend on uranium or fossil fuels. Photovoltaic systems

have many advantages as they do not produce dangerous emissions and they do not cause severe environmental impacts [27]. However, renewable energy sources are costly and have a lower energy density than non-renewable sources.

We have evaluated the following list of the power generation plans show in Table 5.1.

5.2.4 Numerical criteria

Three criteria have been evaluated numerically in this study: energy source, cost and water usage.

Energy Source

Each of the alternatives has a single energy source, which has been obtained from the literature. It has been evaluated with a risk score between 0 (no risk to the environment) and 1 (highest risk to the environment) by a domain expert. The energy sources and their risk scores are shown in Table 5.2. This criterion has to be minimized in order to reduce the impact of the energy source on the environment.

Table 5.2: Values of the Energy Source criterion

Alternatives	Energy Source	Score
NCL	Uranium-U	0.9
NGCC	Shale Gas	0.5
IGCC	Bituminous Coal	0.3
PC	Lignite Coal	0.3
BIO	Energy Crop	0.2
GEO	Geothermal Heat	0.1
WO	Wind	0
SP	Solar Radiation	0
HY	Water	0

Cost

This criterion evaluates the economic cost (£/Mwh) of generating energy for each alternative. Table 5.3 lists the cost for each of them, obtained from the database published in (National Renewable Energy Laboratory,

2015), which includes the costs of operation, fuel and maintenance through
Levelized Cost of Energy (LCOE).

Water usage

In this study we have considered water usage as the water consumption
for the full life cycle stages including the fuel management (its extraction,
processing and transportation) and the power plant life cycle (component
manufacturing, power plant construction, power plant decommissioning
and power plant operation). In addition, we have taken into account that
each power generation technology may use a different cooling system. Then,
we have considered cooling towers for NCL, NGCC, IGCC, PC and BIO,
dry cooling for GEO, and no cooling system in the case of HY, Wind and
SP. The water usage (litres/MWh) reported in Table 5.3 was obtained
from [55], [52].

Table 5.3: Economic and water usage cost of energy generation

Alternatives	Cost (£/Mwh)	Water Usage (litres/MWh)
NCL	100.19	2725.50
NGCC	61.66	794.93
IGCC	131.02	1211.33
PC	115.61	2006.27
BIO	84.78	2093.33
GEO	100.19	3217.60
WO	154.14	3.78
SP	196.68	113.56
HY	77.07	17000.28

5.2.5   Semantic criteria

The semantic criteria considered in this study are Waste By-Products
and Pollution Environmental Damage. The criterion Waste By-Products
describes the different contaminating substances that are produced by
each of the alternative ways of generating energy (e.g. radioactive waste
or CO2). The criterion Pollution Environmental Damage shows different

5.2. Evaluation of energy generation technologies 89

kinds of pollution (e.g. on air, water, soil) and some of its pernicious effects (e.g. acid rain, global warning).

These criteria are multi-valued; therefore each alternative has a list of tags for each of them. These tags are the leaves of the domain ontology shown in Figure 5.6. This ontology was constructed using information from the OUTDO framework and the participation of domain experts [39].
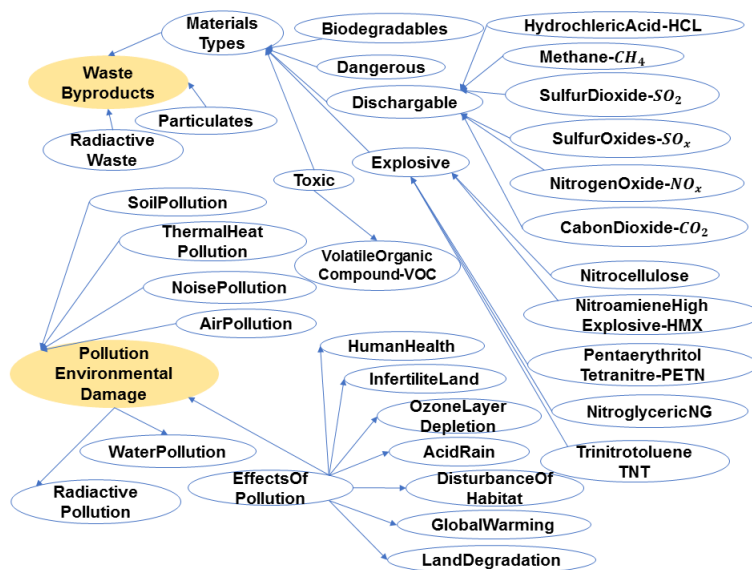
Figure 5.6: Ontology for semantic criteria

In order to evaluate each alternative using these semantic tags, a numerical measure has to be assigned to each tag in the ontology. As proposed before, each leaf a of the ontology stores a function called *TIS(t,a)* defined in chapter 2. It assigns a risk value to each tag t and technology type a depending on the associated numerical measurement $h(a)$. The risk functions have been set using information from the literature and the domain knowledge of the experts on our team.

First, the measurements of each indicator have been extracted from different databases. Table 5.4 shows the emissions of non-renewable (NGCC, IGCC, PC) and biopower technologies for the criterion waste by-products, taken from [14], [46].

Table 5.4: By-product emissions for different kinds of power plants, in g/kWh

Alternatives	CO2	NOX	SOX	CH4	VOC
NGCC	408.7	0.0629	0.0020	0.0079	0.0017
IGCC	716.6	0.2150	0.044		
PC	1003.4	0.94548	2.4057	0.0116	0.0086
BIO	0.0	0.078	0.322		0.070

To assign a TIS to each of the tags for each alternative, the ranges of g/kWh emissions were discretised and a risk score was assigned to each interval with the help of an experts knowledge. The intervals and scores of these by-products are given in Table 5.5.

For the rest of the energy renewable technologies (GEO, WP, SP and HY) the quantities of emissions are negligible, so no risk has been considered (TIS=0, except SO_X and NO_X with 0.1 for GEO). Nuclear plants produce radioactive waste. The TIS score of each semantic attribute for every alternative is shown in Table 5.6.

Table 5.5: Intervals for the waste by-products

CO2 (t)		NOX (t)		SOX (t)		CH4 (t)		VOC (t)	
0-100	0.1	0-0.05	0	0-0.1	0	0-0.002	0		0
100-200	0.2	0.05-0.08	0.1	0.01-0.04	0.1	0.002-0.003	0.1		0.1
200-300	0.3	0.08-0.09	0.2	0.04-0.05	0.2	0.003-0.004	0.2	0.001-0.005	0.2
300-400	0.4	0.09-0.1	0.3	0.3-0.8	0.3	0.004-0.005	0.3	0.005-0.1	0.3
400-500	0.5	0.1-0.2	0.4	0.8-1.00	0.4	0.005-0.006	0.4		0.4
500-600	0.6	0.2-0.3	0.5	1.00-1.20	0.5	0.006-0.007	0.5		0.5
600-700	0.7	0.3-0.4	0.6	1.20-1.50	0.6	0.007-0.008	0.6		0.6
700-800	0.8	0.4-0.5	0.7	1.50-2.00	0.7	0.008-0.02	0.7		0.7
800-1100	0.9	0.5-1	0.8	2.00-2.50	0.8	0.02-0.03	0.8		0.8
1100-1300	1	1-2	0.9	2.5-3.00	0.9	0.03-0.04	0.9		0.9
		2-3	1	3.00-4.00	1	0.04-0.05	1		1

The tags of the pollutants (and their respective TIS) of the Pollution Environmental Damage criterion were assigned by an expert. Again, in renewable technologies we find tags such as Noise Pollution, Land Degradation, Disturbance of Habitat with low risk scores (TIS=0.1), in the cases of WO, SP and HY. Those kinds of power plants generate minimum pollution or environmental damage during their operation processes.

Table 5.6: Tag interest scores (risk) for the values of the semantic criteria for each alternative

Alternative	Waste	TIS	Pollution	TIS
NCL	RadioactiveWaste	1	RadioactivePollution	0.8
			WaterPollution	0.8
NGCC	CO2	0.5	AirPollution	0.4
	SOX	0.1	GlobalWarming	0.3
	NOX	0.1	HumanHealth	0.4
	CH4	0.6		
	VOC	0.2		
IGCC	CO2	0.8	AirPollution	0.6
	SOX	0.2	GlobalWarning	0.7
	NOX	0.5	HumanHealth	0.6
PC	CO2	0.9	AirPollution	0.8
	SOX	0.8	GlobalWarming	0.7
	NOX	0.8		
	CH4	0.7	HumanHealth	0.7
	VOC	0.3	AcidRain	0.8
BIO	SOX	0.3	AirPollution	0.1
	NOX	0.1	GlobalWarming	0.2
	VOC	0.3		
	Biodegradables	0.1		
GEO	SO2	0.1	AirPollution	0.3
	NOX	0.1	GlobalWarming	0.1
	Particulates	0		
	NMVOC	0		
WO	SO2	0	NoisePollution	0.1
	NOX	0		
	Particulates	0		
	NMVOC	0		
SP	SO2	0	LandDegradation	0.1
	NOX	0		
	Particulates	0		
	NMVOC	0		
HY	SO2	0	DisturbanceOfHabitat	0.1
	NOX	0		
	Particulates	0		

The five criteria explained in this section will be used to compare the set of nine power plants. These criteria represent two main issues that are important for the selection of the best plant: amount of resources required (money and water) and environmental impact of the source type (the generated waste and pollution). In the following section, ELECTRE-III-SEM will be used to obtain a ranking of the power plants using this information.

5.2.6   Evaluation using ELECTRE-III-SEM

The parameters of ELECTRE-III-SEM were set in different configurations for this study. Two tests were performed:

- In Test 1, the influence on the ranking of the preference and veto thresholds was analysed in order to study the sensibility of the ranking result to these parameters. For numerical criteria, two scenarios were designed: one with veto power in all criteria and the other without veto power for the numerical features. In the second scenario, numerical criteria may be compensated by good performance in semantic criteria, but not the other way round. This scenario enables the detection of the compensation effects between different types of criteria.

- Test 2 aims to evaluate the degree to which the decision process was affected by the weighting power given to different subsets of criteria. Cases with and without veto power were compared to see how the veto may change the final ranking in criteria with low weight. In this test, two groups of criteria were defined, each one containing criteria of the two types. Therefore, we can analyse the influence of the criteria in the final ranking under different weight conditions, regardless of their type.

In all tests, indifference thresholds q_j were fixed to 0.1 for the semantic criteria and 0 for numerical ones. For the semantic criteria we also fixed μ_j=0.7.

TEST 1: Sensitivity to Preference and Veto

The first test studies the influence of the preference threshold p_j (with values 0.5, 0.3 and 0.1) on semantic criteria. Thus, the veto threshold is

fixed to the maximum possible veto value (which means low discordance effect) for all criteria. This corresponds to $v_j = 0.7$ for semantic data (v^s), and v^n is equal to g_j for numerical criteria. The preference threshold of numerical data is fixed to 20% of the range of g. All criteria have the same weight.

Figure 5.7 shows the rank position (1 being the best) of each alternative. On the left (v^n=70%) we have the three rankings for different values of p_j and the veto power in all criteria. On the right (no v^n) we excluded the discordance step in the numerical criteria (i.e. veto was avoided), so only semantic ones may be in discordance.
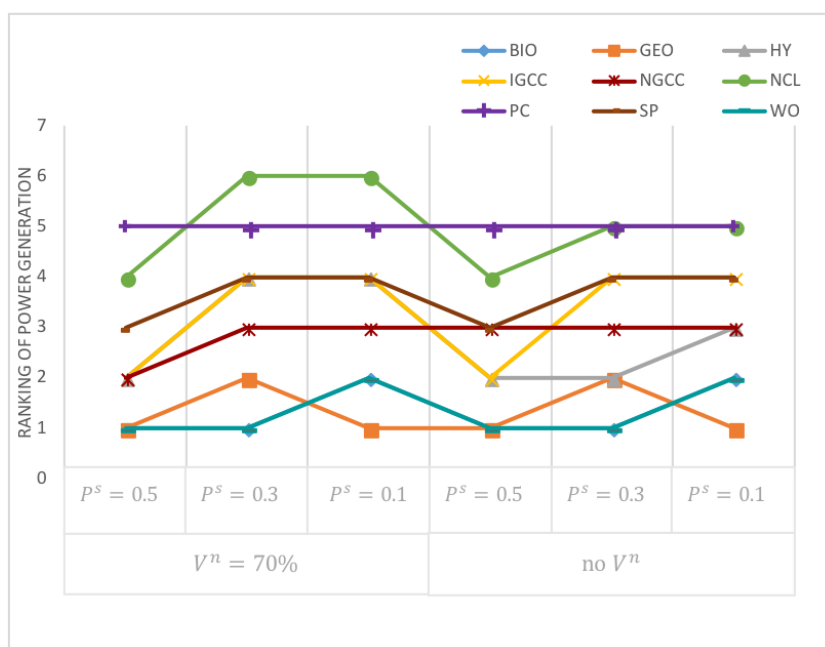
Figure 5.7: Results of Test 1 in two scenarios: with veto in numerical criteria (left), without veto for numerical criteria (right)

Figure 5.7 shows that the best energy sources are always Geothermal, Biopower and Wind. These three energy sources have the best positions for all the values of p. The figure also shows that Hydropower stays in second position for strict models with small p_j. Conversely, the worst power generation plants for the given criteria are Pulverised Coal and Nuclear.

Nuclear power plants are worse than Pulverised Coal when numerical criteria apply veto and the preference threshold is low. Thus, given the criteria used, our method clearly identifies that renewable energy sources outperform the non-renewable ones.

From the previous tables, we can observe that Geothermal is preferable to Biopower regarding the Waste-By-Products criterion, but it is worse in terms of environmental pollution (Pollution-Environmental Damage) in the semantic criteria. Furthermore, for the numerical criteria, Geothermal technology is preferable as Source Energy but it is the most expensive and requires more water. Consequently, depending on the parameters, GEO and BIO exchange the two first positions in the ranking. Hydropower is one of the alternatives that suffers stronger changes of position in the ranking. In addition, Hydropower technology is among the best alternatives for low p_j values (strict configuration) but this alternative is very sensitive to changes in tolerance.

ELECTRE-III-SEM also generates a partial pre-order graph. Figures 5.8 and 5.9 depict the results of Test 1 for the case of p^s=0.5 with veto in all criteria (left graph) and with no veto for water usage and source type (right graph). The graphs show that Wind and Biopower are ranked in the first position because they are not outranked by any other option, but they are incomparable (no one is preferred to the other).

Note that, in the case of no veto (right graph), Geothermal power plants become indifferent to Biopower plants. Wind as a source of energy is expensive but it does not consume water and has very low contamination values. On the other hand, Biopower is a much cheaper energy source and it has few waste by-products, but it uses some water. Therefore, Biopower outperforms Wind in costs but Wind is preferable if water consumption and waste are taken into consideration. Overall, both achieve an equally high performance in comparison with the other alternatives.

We can also see that Hydropower and Solar Power are incomparable to many of the other alternatives. A clear preference is always found between NGCC, IGCC, PC and NCL. Those preference relations may also be useful for the decision maker to make the most convenient selection.

5.2. Evaluation of energy generation technologies 95
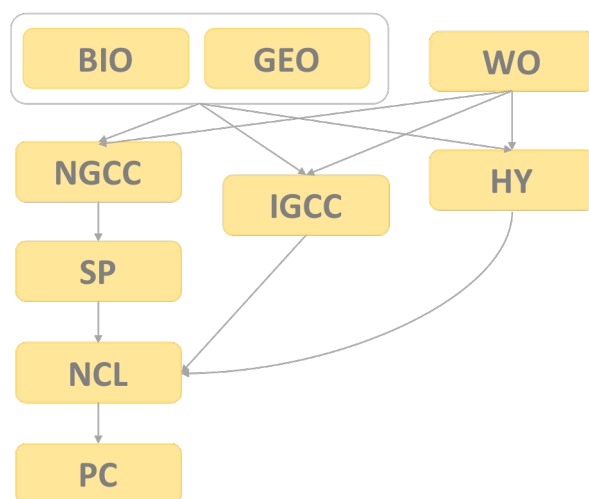
Figure 5.8: Partial pre-orders obtained for Test 1 when using a preference threshold of 0.2 with veto in numerical criteria
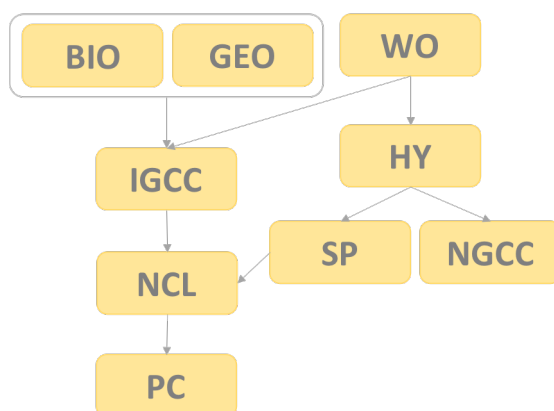
Figure 5.9: Partial pre-orders obtained for Test 1 when using a preference threshold of 0.2 without veto for numerical criteria

TEST 2: Sensitivity to weights on criteria

This test uses the same values as test 1 for the q^n,q^s,v^n,v^s thresholds. A strict preference setting was decided (p^n=20%,p^s=0.5). In this test, we study the results obtained by changing the weight of two groups of criteria - Group A: energy source and water usage and Group B: waste by-products, pollution-environmental damage and cost. A difference of five times is considered, which means that one group of criteria will have five times more voting power than the other group (when calculating concordance). In Case 1, the weight of Group A is 1 and B's weight is 5, whereas the weights are reversed in Case 2. In this test, we compare three situations: first, all criteria have a veto threshold, second, semantic criteria cannot veto (avoid sem) and third, numerical criteria cannot veto (avoid num).
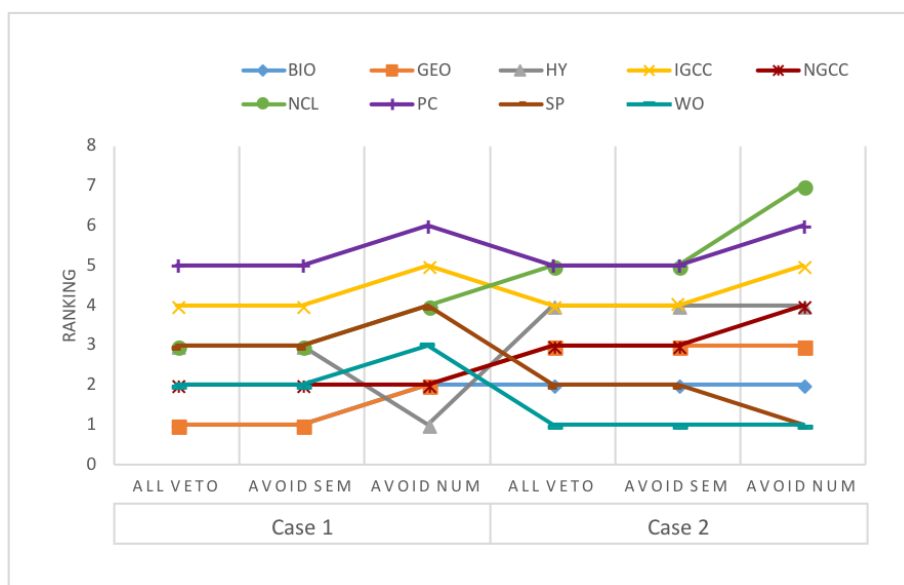
Figure 5.10: Results for Test 2 considering different weights on two groups of criteria

From Figure 5.10, we can observe that when semantic information and cost have a larger weight (Case 1), the best options are Wind, Hydropower, Geothermal and NGCC. When we decrease the importance of these criteria and increase the weight of water usage and energy source (Case 2), the best options are Wind, Solar Photovoltaic, Biopower and NGCC.

In general, for non-renewable technologies the ranking is almost the same (they are in the worst positions, except NGCC). It can be observed that, in the first case, Wind descends positions when numerical criteria do not veto the outranking relation. Wind is one of the best in terms of energy source and water usage, but in Case 1 these two criteria have a low weight, and therefore the only way to influence the ranking result is by vetoing Group A. When we do not allow the veto power for those criteria, they are almost neglected in the calculation of the ranking, thus Wind goes from the first position to the third.

In Case 2, we can see that Solar Photovoltaic technology is considered among the best options because we are giving more importance to energy sources and water usage (Group A). Significant differences in rank positions can be found when changing the balance of the criteria in favour of one or another set (e.g. Hydropower, Geothermal and Solar). For instance, Hydropower moves from the first position in Case 1 to the seventh position in Case 2. However, we notice that non-renewable technologies are always in the last positions. According to [18], wind power is growing due to the increasing prevalence of wind-generated electricity in many countries. These results coincide with the recommendation to use renewable power technologies.

The analysis done in this section is highly sensitive to each country, both in the semantic and quantitative variables. For example, costs may be different in other locations. Regarding the semantic criteria, although the tags will be the same all around the world (because pollutants and waste depends on the technology type and not on the location), the subjective evaluation of risks may be different depending on the conditions of each place. Moreover, the parameters used in the model (thresholds, veto power, weights) also greatly depend on the experts requirements.

5.3   A recommender system for tourists visiting Costa Daurada & Terres de l'Ebre

This section describes a case study related to the recommendation of touristic activities. More concretely, we compare the results of the ELECTRE-SEM method proposed in this thesis with the ones produced by a recommender system called SigTur. This recommender system was

developed at PCTTO (Parc Científic i Tecnològic de Turisme i Oci, Scientific and Technological Park for Tourism and Leisure) in collaboration with other members of my research group ITAKA at URV [10].

This recommender system has an extensive database with thousands of touristic activities from the Catalan region known as "Costa Daurada and Terres de l'Ebre", which is located at the south of Catalonia. The system includes different types of cultural and leisure activities that can be done in this area. It also uses an ontology-based user profile and an ELECTRE-based ranking procedure, but the methods used for managing the tag interest scores are completely different from the ones presented in this thesis. In SigTur the scores of the tags are aggregated into a unique overall value for each alternative. This aggregation is done with the *Ordered Weighted Average* operator (OWA). The result is a numerical criterion that gives a quantitative score to each activity according to the average of its tag scores. Then, as it is a numerical criterion, the classic ELECTRE procedures (concordance and discordance) can be applied to build the credibility matrix of the outranking relation. This matrix is exploited using the Net Flow Score ranking technique (section 4.2.2).

To make a comparison of ELECTRE-SEM and the SigTur system, we will consider two different ways of calculating the concordance and discordance indices from the tag interest scores, while keeping the rest of the process the same (i.e. preference elicitation and ranking with NFS):

A. Transform the semantic criterion into a numerical one using the Ordered Weighted Average operator [97], [10].

B. Keep the semantic criterion and use the ELECTRE-SEM concordance and discordance indices based on the Semantic Win Rate (SWR) defined in this dissertation [53].

The first approach summarizes a list of tag interest scores into a single overall score, which may produce a loss of information (e.g. an average score of 0.5 might be obtained by aggregating several medium scores or by merging some high scores with some low ones). The tools developed in this PhD thesis provide a new way of handling semantic criteria that avoids the transformation of the type of data. The aim of this section is to analyze the differences obtained in the ranking of the alternatives depending on the method employed to manage the semantic criterion.

5.3.1   The SigTur recommender system

This section presents the SigTur method for rating and ranking alternatives, which are tourism and leisure activities. In SigTur each activity is tagged with one or more ontology concepts, which are leaves (or low level nodes) in the ontological hierarchy. For each user there is a personal ontology that contains a preference degree for each of the classes, depending on the explicit and implicit information provided by the user. As the user knwoledge is mainly obtained in an implicit way, a confidence score is also associated to each of the preference scores. Both values are used to decide which activities are recommended to the user.

An important element of SigTur is the use of a domain ontology to guide the recommendation process, which permits to make inferences about the correspondence between the characteristics of an activity and a certain user profile. SigTur makes a knowledge-level analysis of the user preferences, including processes that make bottom-up and top-down propagation of the preferences over the concepts of the ontology. This information is very useful in order to take the final decision of which activities to show to the user.

In [10] an ad-hoc domain ontology was defined, following the principles of the thesaurus of the World Tourism Organization but adjusting it to the specificity of the territory. The ontology, which was manually created, represents up to 203 connected concepts in 5 hierarchy levels. The ontology is structured around eight main concepts that constitute the first level of the hierarchy: Events, Nature, Culture, Leisure, Sports, Towns, Routes and ViewPoints. The last three classes are considered traversal concepts, since they share children nodes with other main classes (e.g., Routes and Nature are both super-classes of the NatureRoutes class). The rest of the concepts in the ontology are connected via is-a (subclass) relationships with these main classes. The ontology is not a pure taxonomy, as it contains multi-inheritance between concepts (e.g. EthnographicMuseum is a subclass of both Museum and Traditional).

The decisions about which concepts and relationships should be represented in the ontology have been taken by a committee of experts in the Tourism domain from the Science & Technology Park for Tourism and Leisure. The level of detail in each part of the ontology depends on the set of activities available in the particular geographical area of

interest. For example, there is a deep level of detail about concepts related to Wine due to the importance of enotourism in the region. In any case, the ontology could be easily extended with more concepts if it were necessary. For instance, this ontology could be customized to another region where winter sports were relevant, by adding a new concept called WinterSports (with its appropriate subclasses) and putting it as a subclass of the NonAquaticSports concept.

A first step to develop SigTur was to collect various data sets of tourism resources (leisure activities, cultural heritage, natural spaces, sport activities, routes and events) of the Tarragona province to build a GIS database. This information was spread in different government administrations; therefore, the first task was to request these data sets. Most of them were obtained from Diputació de Tarragona, although an important part was provided by Generalitat de Catalunya.

The activities of the GIS database of SigTur are grouped into six categories: leisure, sports, culture, nature, events and routes. The last two play a cross-cutting role, since they can be related to any of the other categories. Items associated to the ontology concepts towns and viewpoints are always stored in one of these categories (for instance, an item tagged as CultureViewpoints or TraditionalTowns would be stored in the Culture category). Leisure contains five entities (equivalent to tables or map layers): beaches, theme parks, spa centers, shopping areas and nightlife areas. The data of these entities have been added to the database with special care, performing an exhaustive documentation task, since they are the main tourist attractions in Tarragona. Sports have been classified in two subcategories: aquatic and non-aquatic. Culture includes two entities: cultural heritage assets and museums. They are stored in different tables since the structure of their information is relatively different. Nature contains two entities: natural spaces, which encompass all the natural spaces protected by law, and the recreational areas contained within these spaces. Events include temporary activities (such as fairs, festivals, traditional celebrations, and so on) that can be programmed throughout the year in any of the other categories. Finally, routes include three entities that can also be related to the other categories: walking routes, biking routes and driving routes. Currently, the GIS database contains over a thousand resources. Nevertheless, there is still a considerable ongoing

5.3. A recommender system for tourists visiting Costa Daurada & Terres de l'Ebre

work on adding new resources and updating the existing ones. In any case, the GIS database has been designed in order to easily support these future additions and updates.

The SigTur recommender system manages a user profile that is composed by two parts: (1) a static one, which is a vector with demographic and travel information and (2) a dynamic one, represented with an instantiation of the Tourism ontology, which contains the users degree of interest on each type of activity. For instance, if the current user of the system likes visiting museums and is especially interested in wines, the concepts Culture, Museums and particularly WineMuseums will have a higher degree of preference than others. This part of the profile is updated when new knowledge is obtained from the user.

The first task of a user in the system is to complete a form, which is used to create the initial profile. The main goal is to obtain as much information as possible with a small number of questions. The Tourism partners of the SigTur project elaborated a survey questionnaire to discover the most common travel motivations of the tourists that visit the Tarragona region. From a statistical analysis of thousands of surveys, it was discovered that the main motivations (sorted in order of importance) were the following: beach, shopping, relaxation, leisure, culture, nature, gastronomy, sports and shows/events. Each of these motivations corresponds to a concept stored as a class in the Tourist ontology that may either be at the top level of the ontology, such as Leisure or Culture, or at lower levels, such as Beaches or Shopping. Even though the concepts Beaches and Shopping are children of the Leisure concept, we decided to ask independently about the three motivations due to their importance on the survey analysis. These values are stored in the ontology of the user to initialize his/her profile.

The data needed to initialize the demographic and travel information of the user is obtained also with a form presented to the user at the beginning of the session (see Figure 5.11). These data include information about the country of origin of the user, other people the user travels with, the location of the accommodation, the type of accommodation, an initial estimation of the budget, and the travel dates. Some of those variables are used to filter the results before they are shown to the user (travel dates) or to locate the recommendations into a given geographical area (near the chosen destination) [10].
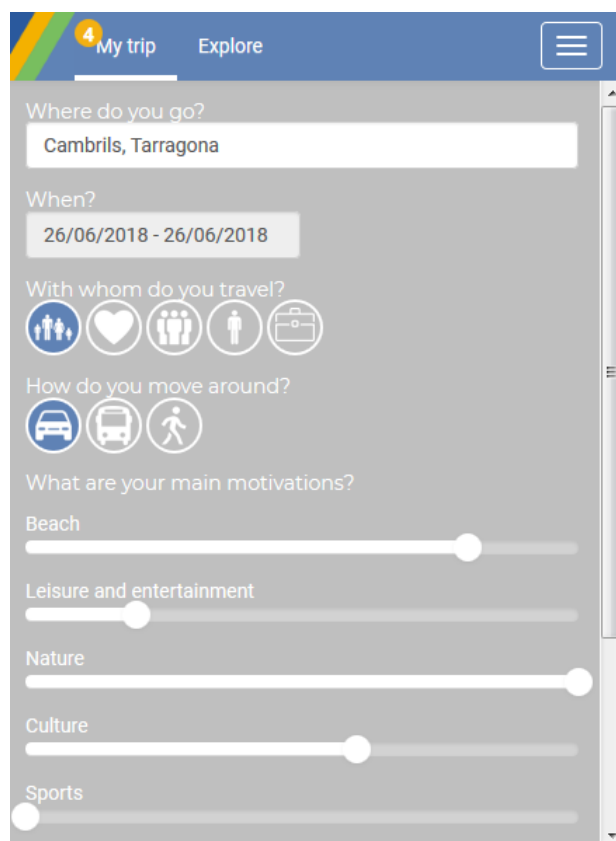
Figure 5.11: SigTur interface to introduce the user's main preferences

Apart from the explicit information given at the beginning of the session by the user, the system is able to obtain explicit information from the evaluations that users make on the activities they have already visited, in which they express their degree of satisfaction. Users may rate activities with an integer value between 1 and 5, where 5 corresponds to the best. The system also takes into account the actions performed by the user during his/her interaction with the system, in order to improve the information about the users preferences and its recommendations. Once the user obtains a list of recommendations he/she can make several actions on the proposed activities. The system is able to infer the users interests by capturing and

analyzing these actions. This process is very useful to adapt dynamically and automatically the user profile and make more precise the degree of interest of the user on each kind of activity during the recommendation session (see Figure 5.12). For example, the user may select those activities he/she is interested in and add them to a travel plan. Other actions the user is able to make on activities are to request more detailed information on a specific event, to ask for activities geographically close to the currently selected one or to obtain activities that are thematically similar to the current one.
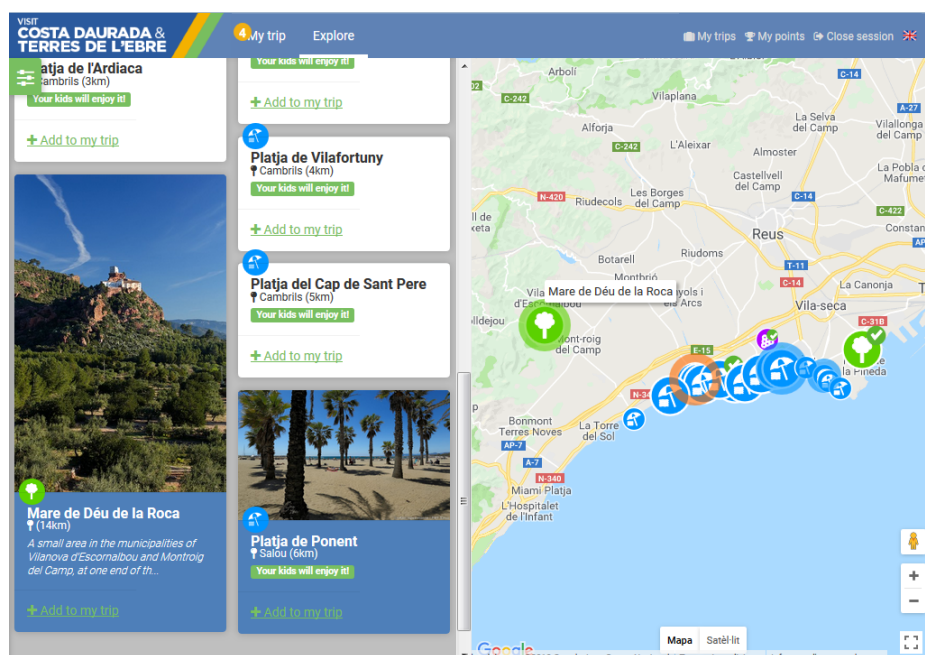
Figure 5.12: SigTur window to display the recommended activities

The information obtained from the user actions is mapped into preferences related to the concepts associated to the manipulated activities, which are nodes in the lowest levels of the ontology. A spreading algorithm has been used to propagate the preference values of the ontology nodes to their ancestors. This process, has two steps: upwards propagation (in which the interests on the ancestors of the modified leaves are updated)

and downwards propagation (in which the preference and confidence on other descendants of these ancestors are also updated).

In order to perform a user-based collaborative recommendation it is necessary to have a way to compare two users, which gives us an estimation of their similarity. This measure can then be used to automatically build groups of similar users. In [65], a similarity measure based on demographic and motivational attributes was used for building clusters of similar users. So, finally, the recommendation is based on several criteria, not only the semantic tags related to the type of touristic activity (TTAG), but also the collaborative information (COLAB), the distance (DIST) or the price (COST).

As we cannot change SigTur, in order to make a comparison with ELECTRE-SEM we have defined the following procedure, which has been run in the ELECTRE software package:

1. Select a subset of criteria used in SigTur: distance, cost, collaborative score and semantic tags.

2. Select a subset of 70 diverse touristic activities.

3. Define a user profile, with a fixed destination and budget. Then, make 4 profiles with different semantic preferences on the tags of the alternatives.

4. Run SigTur with these profiles and extract the TIS stored by the system in each concept and the collaborative value given to each alternative for each profile.

5. Create two input data matrices for the ELECTRE software package with these data: distance, price, collaborative score and with the OWA result of the TIS (version 1 - OWA) or with the semantic multi-valued criterion directly (version 2 - SWR).

6. Run the ELECTRE software package with the two versions using NFS and then compare the results.

These steps are detailed in the following sections.

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

5.3. A recommender system for tourists visiting Costa Daurada & Terres de l'Ebre 105

5.3.2   User profiles for testing

To make a comparison of the method proposed in this thesis with the SigTur recommender system, we have considered a set of user profiles predefined by the tourist experts. We have based the definition of the profiles on the stereotypes that were defined in the doctoral thesis *"Provision of personalised information about tourist activities"* of Joan Borràs (page 162) [9], Technical Head of the Scientific and Technological Park for Tourism and Leisure and main developer of SigTur. More concretely, for the illustrative purposes of this case study the following stereotype was selected:

A group of young Spanish friends that visit Tarragona, who are 20-25 years old and stay in a 3-stars hotel.

Consider this kind of users and their characteristics, we have defined four different profiles that are represented by different preference scores for the types of activities. They are also based on the stereotypes of this kind of visitors in this area.

- Profile 1: Their main motivations are going to the beach, make leisure and gastronomy-related activities, and attend events. They also have a small interest in culture or shopping, but they don't enjoy relaxation activities and sports.

- Profile2: In this second profile, the visitors prefer to spend time on relaxation activities, Nature-related tours and leisure activities. They are less interested in sports, gastronomy and events. They don't have any interest at all in cultural activities, shopping or going to the beach.

- Profile 3: This kind of users is mainly interested in cultural and gastronomic activities and events. They may also enjoy shopping and relaxation, but they are definitely not interested in sports, Nature and going to the beach.

- Profile 4: In this case the main travel motivations are going to the beach, shopping and sports, followed by Nature, relaxation and events. They do not enjoy cultural and leisure activities.

These general interests have been translated into the percentages given in Table 5.7, which have been used to define the user's interest on each of the concepts appearing in the multi-valued criterion TTAGS.

Table 5.7: Travel motivations

Topics	Profile 1	Profile 2	Profile 3	Profile 4
Beach	81 %	5 %	5 %	80 %
Shopping	15 %	1 %	18 %	58 %
Relaxation	6 %	90 %	20 %	33 %
Leisure and entertainment	45 %	52 %	13 %	7 %
Culture	17 %	0 %	76 %	6 %
Nature	11 %	85 %	5 %	38 %
Gastronomy	41 %	20 %	55 %	22 %
Sports	8 %	23 %	0 %	54 %
Shows and events	38 %	18 %	45 %	30 %

In order to apply the ELECTRE method it is necessary to set the values of the parameters of each criteria (see Table 5.8). First of all, the same *weight* has been given to all criteria (25% each). The *Type* row indicates the types of treatment of the criteria (semantic or numerical). The *direction* indicates if the value of the criterion has to be maximinized or minimized. Finally, the last rows correspond to the thresholds. In particular, the *indifference*, *preference* and *veto* thresholds determine the way in which the performance values of different activities will be compared. The thresholds for TTAGS are different depending on the type of treatment that is done to this criterion (OWA or SWR), since they don't have the same meaning, as we have seen in the previous chapters. The thresholds of the numerical attributes do not change in both tests, as these criteria are treated in the same way.

5.3. A recommender system for tourists visiting Costa Daurada & Terres de l'Ebre 107

Table 5.8: Parameters of the ELECTRE method

CRITERION	SWR TTAGS	OWA TTAGS	DIST (KM)	COST ()	COLAB
Type	sem	num	num	num	num
Weight	0,25	0,25	0,25	0,25	0,25
Direction	max	max	min	min	max
Indifference	0,05	0,05	1	5	0,1
Preference	0,5	0,15	3	10	0,2
Veto	0,3	0,3	6	15	0,5
Min SWR	0,7	NA	NA	NA	NA

The indifference of the TTAGS criterion is the same because it directly refers to the Tag Interest Score (TIS). However, for the preference and veto thresholds the value is different because in the SWR method it refers to the Semantic Win Rate, $SWR(a, b)$, while in the OWA method it refers to the overall numerical TIS obtained after the aggregation. The OWA thresholds have been determined after studying the distribution of preference scores on the 70 activities of this case study, shown in Figures 5.13, 5.14, 5.15 and 5.16. For SWR, preference and veto values are defined in relation to *Min SWR*, which is the minimum value of SWR that leads to full concordance, in this case 0.7. Then the minimum value for weak concordance was set to 0.5 and the maximum value of SWR that leads to a full discordance is 0.3 (which means that between 0.3 and 0.5 we consider weak discordance).
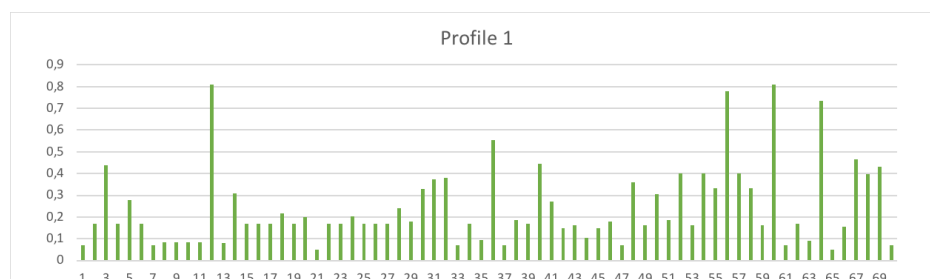
Figure 5.13: Preference on each activity - OWA Profile 1
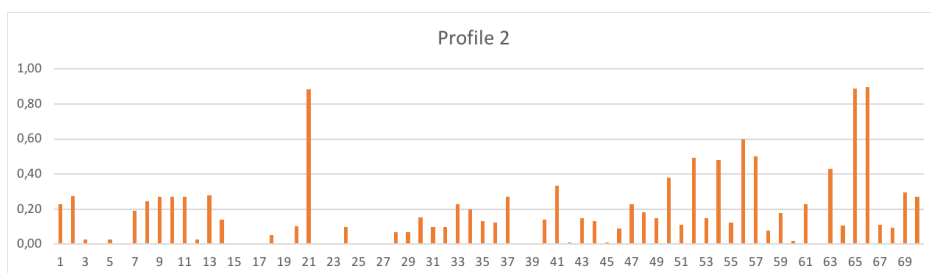
Figure 5.14: Preference on each activity - OWA Profile 2
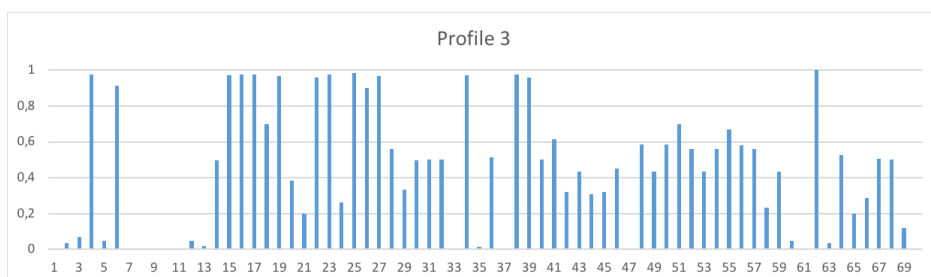
Figure 5.15: Preference on each activity - OWA Profile 3
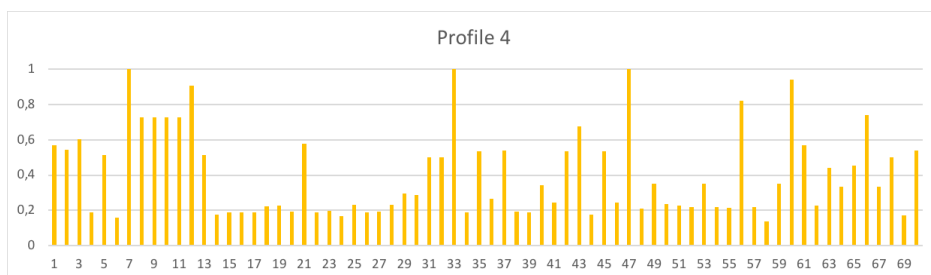
Figure 5.16: Preference on each activity - OWA Profile 4

5.3.3 Execution of the recommender system with the two methods

The first execution was done with the numerical version of the Touristic Tags criterion, with the values obtained using the OWA aggregation operator with a conjunctive set of weights (with orness 0.2). This corresponds to the current implementation in SigTur.
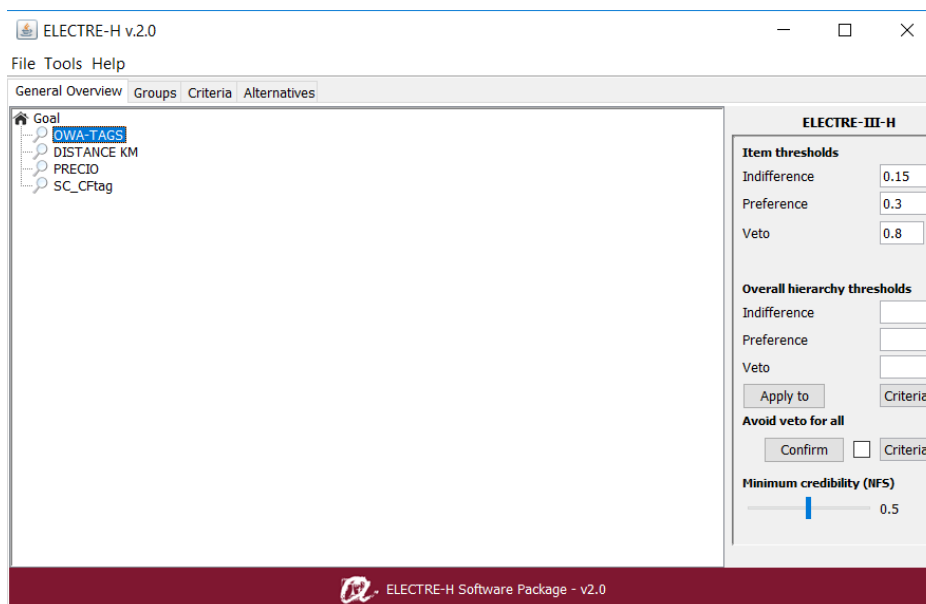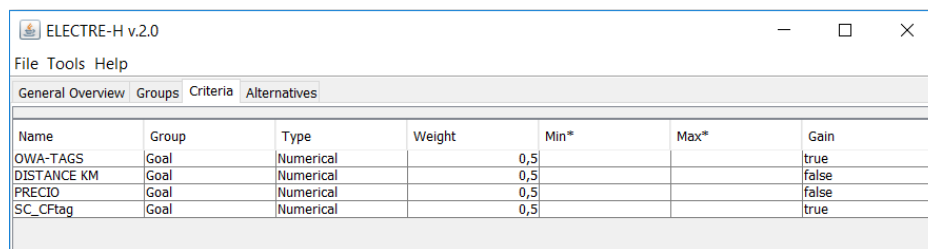
Figure 5.17: ELECTRE software package main window - OWA test
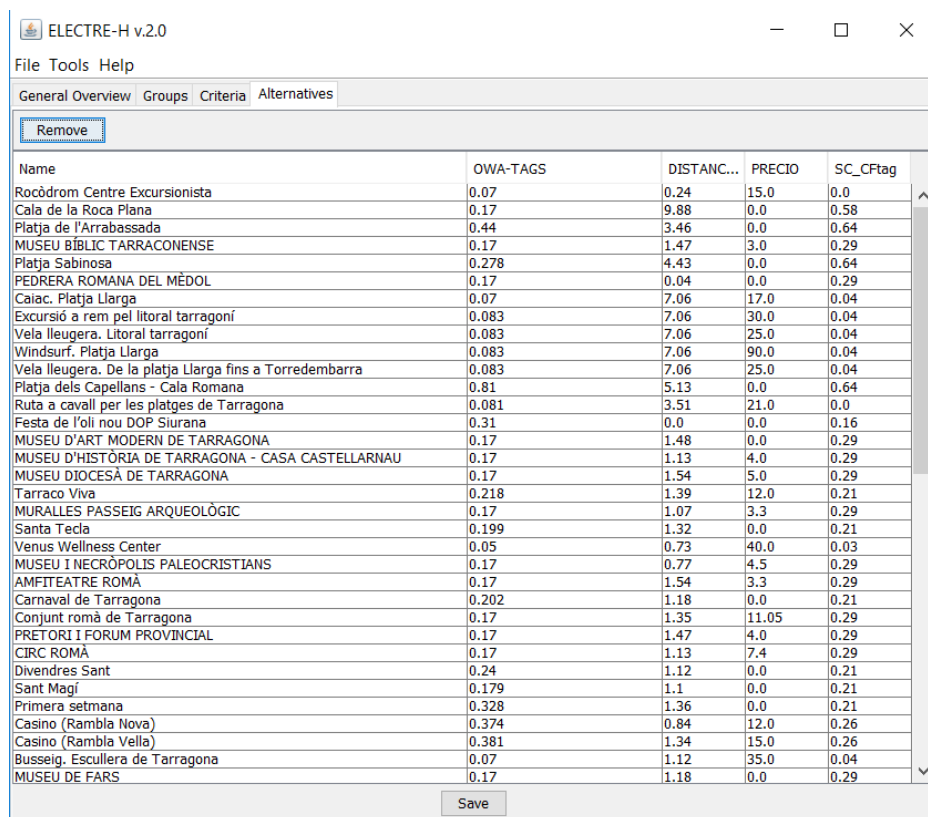
In the interface of the software package (Figure 5.17) it may be seen that the four criteria are introduced at the same level, with their corresponding thresholds (right hand side). The NFS ranking method has been executed with a minimum credibility level of 0.5. The details of the criteria can be seen in the "Criteria"' tab of the software (see Figure 5.18). They are all defined as numerical and with equal weights. Its direction (gain or cost) is also indicated. A part of the performance matrix with the corresponding OWA score is shown in Figure 5.19. In this case study we have employed a set of 70 touristic activities of Costa Daurada and Terres de l'Ebre.

Data must be introduced in a different way for the test that uses the

Figure 5.18: ELECTRE software package: List of criteria - OWA test

Figure 5.19: ELECTRE software package: Performance matrix - OWA test

Figure 5.20: ELECTRE software package main window - SWR test

Semantic Win Ratio and the new semantic concordance and discordance indices defined in this dissertation. Figure 5.20 shows a similar tree, but the first criterion is now treated in a semantic way, as seen in Figure 5.21. The system also allows to see the details of the ontology-based user profile for the semantic criterion. Figure 5.22 displays a portion of the ontology related to sport concepts and their tag interest scores. Some tags do not have any score because the user has still not given a value for them. In this case, these tags do not appear in any of the activities in the dataset. In this experiment we have the TIS for all the tags that appear in the activities in the dataset. If they were not known, the procedure explained in the first part of the thesis could be used to infer the missing values. Finally, Figure 5.23 shows a portion of the performance matrix with the details of the semantic criterion. It can be seen here that all the tags have a valid numeric interest score.

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

112   **Chapter 5.   Tools and applications**

Figure 5.21: ELECTRE software package: List of criteria - SWR test
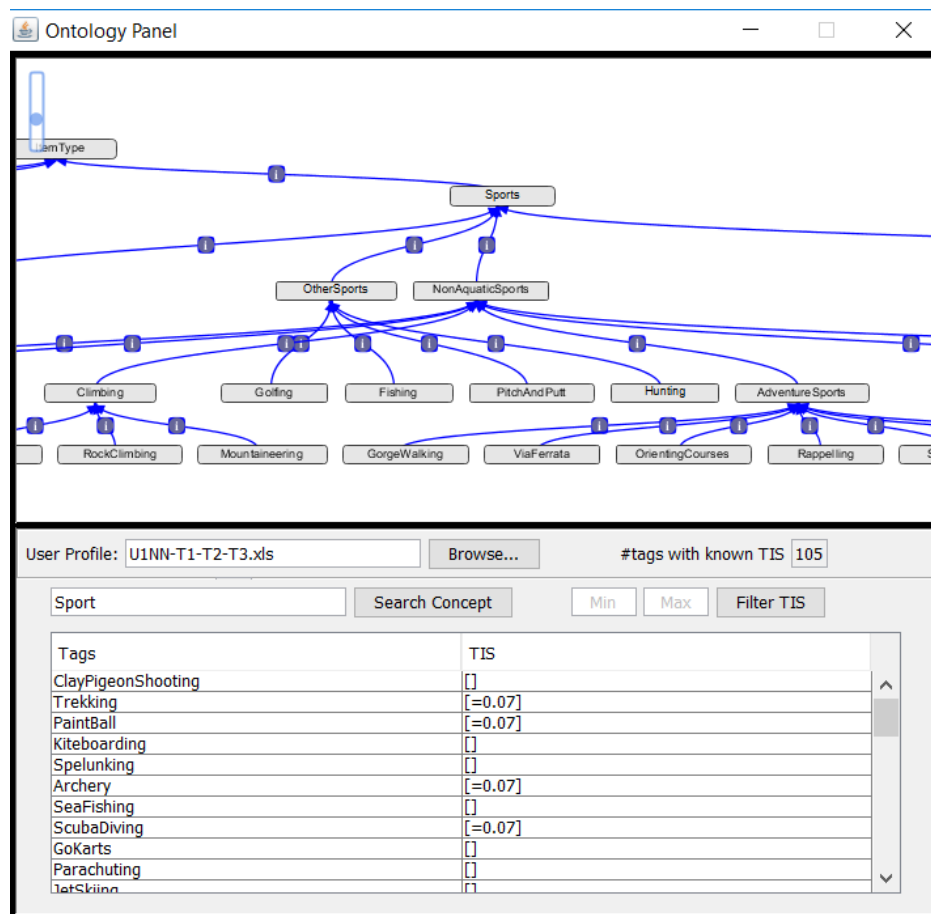
Figure 5.22: ELECTRE software package: View of ontology and TIS

5.3. A recommender system for tourists visiting Costa Daurada & Terres de l'Ebre   113

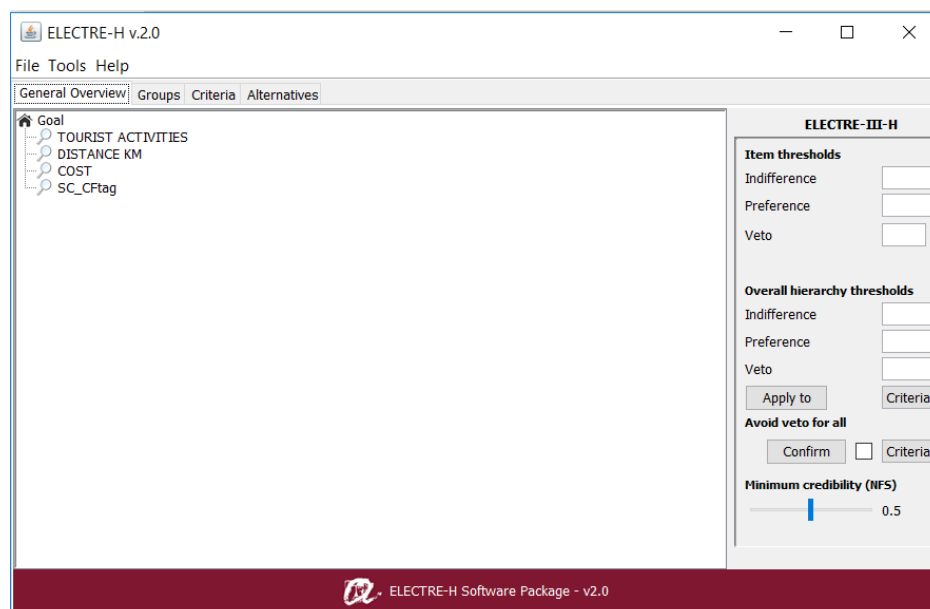| Name	TOURIST ACTIVITIES	DISTANC...	COST	SC_CFtag
Rocòdrom...	{ClimbingWall=0.07, RockClimbing=0.07}	0.24	15.0	0.0
Cala de la...	{NudismBeaches=0.81, NaturalSpaces=0.1, InlandWatersRoutes=0.1, CoastalRoutes=0.1, ViaFerrata=0.07, Coves=0.81}	9.88	0.0	0.58
Platja de l'...	{BeachPicnic=0.81, FamilyBeaches=0.81, RefreshmentStall=0.4, NormalBeaches=0.81, PicnicAreas=0.1}	3.46	0.0	0.64
MUSEU Bĺ...	{SacredArtMuseums=0.17, Roman=0.17}	1.47	3.0	0.29
Platja Sab...	{NudismBeaches=0.81, CoastalRoutes=0.1}	4.43	0.0	0.64
PEDRERA ...	{Ruins=0.17, CultureRoutes=0.17, Roman=0.17, HumanHeritage=0.17, HistoryMuseums=0.17, ArcheologyMuseums=0.17}	0.04	0.0	0.29
Caiac. Plat...	{Canoeing=0.07, Kayaking=0.07}	7.06	17.0	0.04
Excursió a...	{Boating=0.07, InlandWatersRoutes=0.1, Rowing=0.07, Kayaking=0.07}	7.06	30.0	0.04
Vela lleug...	{Boating=0.07, Sail=0.07, InlandWatersRoutes=0.1, Windsurfing=0.07}	7.06	25.0	0.04
Windsurf. ...	{InlandWatersRoutes=0.1, Windsurfing=0.07, Wakeboarding=0.07, Kitesurfing=0.07}	7.06	90.0	0.04
Vela lleug...	{Boating=0.07, Sail=0.07, InlandWatersRoutes=0.1, Windsurfing=0.07}	7.06	25.0	0.04
Platja dels...	{BeachPicnic=0.81, NormalBeaches=0.81}	5.13	0.0	0.64
Ruta a cav...	{Trekking=0.07, NaturalSpaces=0.1, HorseRiding=0.07, CoastalRoutes=0.1, CoastalTowns=0.1}	3.51	21.0	0.0
Festa de l'...	{PopularCelebrations=0.32, TraditionalCousine=0.33, OilRoutes=0.3, WineFairs=0.35, WineFestivals=0.3}	0.0	0.0	0.16
MUSEU D'...	{PaintingMuseums=0.17, ArtMuseums=0.17}	1.48	0.0	0.29
MUSEU D'...	{InterpretationCenter=0.17, Roman=0.17, HistoryMuseums=0.17}	1.13	4.0	0.29
MUSEU DI...	{Roman=0.17, ArtMuseums=0.17}	1.54	5.0	0.29
Tarraco Vi...	{Medieval=0.17, GastronomyFestivals=0.39, TraditionalCelebrations=0.35, Roman=0.17}	1.39	12.0	0.21
MURALLE...	{Roman=0.17, HumanHeritage=0.17, HistoryMuseums=0.17, Walls=0.17}	1.07	3.3	0.29
Santa Tecla	{PopularCelebrations=0.32, MusicFestivals=0.37, TraditionalCelebrations=0.35, LocalProducts=0.15, ArtsAndCraftsEvents=0.29, WineAtmosph...	1.32	0.0	0.21
Venus We...	{HealthResorts=0.05, SpaResorts=0.05, HydroFacilities=0.05}	0.73	40.0	0.03
MUSEU I ...	{Roman=0.17, HumanHeritage=0.17, ArcheologyMuseums=0.17}	0.77	4.5	0.29
AMFITEAT...	{Ruins=0.17, Amphitheater=0.17, Roman=0.17, HumanHeritage=0.17, HistoryMuseums=0.17, Romanesque=0.17}	1.54	3.3	0.29
Carnaval ...	{DanceFestivals=0.37, MusicFestivals=0.37, TraditionalCelebrations=0.35, BigGroupsAtmosphere=0.0}	1.18	0.0	0.21
Conjunt ro...	{Aqueduct=0.17, Amphitheater=0.17, Circus=0.17, Tower=0.17, Roman=0.17, HumanHeritage=0.17, Walls=0.17}	1.35	11.05	0.29
PRETORI I...	{Palace=0.17, Roman=0.17, HumanHeritage=0.17, HistoryMuseums=0.17, Gothic=0.17}	1.47	4.0	0.29
CIRC ROMÀ	{Circus=0.17, Roman=0.17, HumanHeritage=0.17, HistoryMuseums=0.17}	1.13	7.4	0.29
Divendres ...	{PopularCelebrations=0.32, TraditionalCelebrations=0.35, Cathedral=0.17}	1.12	0.0	0.21
Sant Magí ...	{RuralAtmosphere=0.1, PopularCelebrations=0.32, TraditionalCelebrations=0.35, Chapel=0.17}	1.1	0.0	0.21
Primera s...	{PopularCelebrations=0.32, TraditionalCelebrations=0.35}	1.36	0.0	0.21
Casino (R...	{GameRoom=0.36, Discos=0.39, Bars=0.4}	0.84	12.0	0.26
Casino (R...	{BallRoom=0.43, GameRoom=0.36, Bars=0.4}	1.34	15.0	0.26
Busseig. E...	{Snorkelling=0.07, ScubaDiving=0.07}	1.12	35.0	0.04
MUSEU DE...	{SeaportMuseums=0.17}	1.18	0.0	0.29
Creuers in...	{Boating=0.07, CoastalRoutes=0.1}	1.25	190.0	0.04
Xiringuito ...	{Pubs=0.41, TastingCousing=1.0, TapasCousine=1.0, RefreshmentStall=0.4}	4.47	0.0	0.26
Club de G...	{PitchAndPutt=0.07, Golfing=0.07}	7.69	40.0	0.06
MUSEU DE...	{HistoryMuseums=0.17, SeaportMuseums=0.17, HistoricBuilding=0.31}	1.94	0.0	0.29
MNAT VIL...	{Roman=0.17, HumanHeritage=0.17, ArcheologyMuseums=0.17}	4.81	2.5	0.29
Casino Sa...	{RegionalCousine=1.0, GameRoom=0.36, Bars=0.4}	4.67	35.0	0.26
Dalmau H...	{AgroShop=0.4, Vineyards=0.23}	2.3	45.0	0.0
El Corte In...	{ShoppingCenter=0.15, ShoppingArea=0.15}	0.46	0.0	0.7
Carrer Ma...	{LocalProducts=0.15, LocalMarket=0.2}	1.3	0.0	0.7
Mercat Ce...	{UniqueBuilding=0.31, OutstandingProduct=0.0, LocalProducts=0.15, LocalMarket=0.2}	2.24	0.0	0.7
Parc Central	{ShoppingCenter=0.15, ShoppingArea=0.15}	0.85	0.0	0.5
Via T	{FoodAreas=0.4, LocalProducts=0.15, ShoppingArea=0.15}	0.97	0.0	0.5
Submarini...	{Snorkelling=0.07, ScubaDiving=0.07}	3.09	25.0	0.04
Bodegas ...	{WineShops=0.4, CellarShop=0.4, ModernistCellar=0.31, Cooperatives=0.4}	9.6	35.0	0.0
Mercadet ...	{LocalProducts=0.15, LocalMarket=0.2}	5.11	0.0	0.7
La Vinícol...	{WineShops=0.4, Vineyards=0.23, WineRoutes=0.4}	7.61	55.0	0.0
Mercadet ...	{LocalProducts=0.15, ArtsAndCraftsEvents=0.29}	3.11	0.0	0.7
Celler Ca...	{AgroShop=0.4, WineTasting=0.4, WineRoutes=0.4}	9.97	24.9	0.0

Save

Figure 5.23: ELECTRE software package: Performance matrix - SWR test

5.3.4   Results

This section shows the recommended activities for each user profile in both methods (OWA -actual version of SigTur- and SWR -proposed in this dissertation-). We will focus on the top 20 results, as they correspond to the activities that would be initially shown to the user. For each profile we have obtained the net flow scores to construct the ranking of the alternatives. We have tested the following three situations:

- Veto in all the criteria.

- Without veto in the semantic criterion (touristic tags).

- Without veto in the numerical criteria (distance, cost and collaborative).

Although the preferred configuration is the first one (using the veto power in all criteria), the other two tests have been made to study the effect of the semantic criterion on the ranking in both methods. When the semantic criterion cannot veto an outranking relation, the numerical criteria will mainly determine the ranking. On the contrary, when numerical criteria don't have the possibility to veto an outranking relation, the semantic criterion will be more relevant in the determination of the final result.

Results for PROFILE 1

In this profile, the group of youth are mainly interested in visiting the beach, followed by doing leisure activities and attending events as well as in gastronomy. They also have a small interest in culture or shopping. They do not like relaxation activities nor sports.

The first table of results (Table 5.9) shows the top 20 best options in each method, ordered by the inferred user's preference. We can see that the first 4 positions are quite similar, just with small rank reversals, due to the fact that SWR prioritizes alternatives with a larger set of tags with high TIS (i.e. Platja de l'Arrabassada is better than Platja Cala Romana because it has more descriptors). Two beaches are in the top positions in both cases. The following positions in the SWR-based ranking show the Casino, the Carnaval cellebration and another beach, which fit

nicely with the user's interest in Gastronomy, Leisure and Beach. However, in positions 4, 5, 6, 7, 9 and 11 of the OWA ranking there are mainly markets and shopping centers, in which this kind of users were not very interested (15%). The OWA ranking also shows several cultural activities such as museums (only 17% interest), which do not appear in our new ranking. Moreover, the new SWR ordering shows many more events (38% interest) than the OWA one (e.g. in positions 6, 8, 10, 11 and 14). Some important differences in the ordering of events may be noticed, for example the following:

- Primera Setmana: SWR position is 8, OWA position is 14.

- Festa Oli DOP Siurana: SWR position is 10, OWA position is 15.

- Santa Tecla: SWR position is 11, OWA position is 32 (not shown in the top 20).

- Divendres Sant: SWR position is 14, OWA position is 30 (not shown in the top 20).

Thus, in general, it seems that the SWR ranking fits better with the kind of activities preferred by this kind of users. The difference in the NFS value in the two rankings indicates that OWA produces much more ties in the evaluation of the pairs of alternatives, because at position 20 a museum (not a very good for this user profile) still receives an score of 23. Thus, this option has a difference of 23 in strength-weakness out of 70, that is 46-24, meaning that it is 46 times equal or better -at least as good as - the rest of alternatives. As we know that a museum has not a high TIS, the conclusion is that this strength may probably correspond to the similarity of the OWA scores obtained from the semantic criterion. In fact, it was shown in Table 5.13 that there are many activities with a low interest score for this kind of users.

In Table 5.10, we can see the differences in the recommended activities when some criteria are not allowed to use the veto power. In all the options there are two beaches at the top positions, which are activities related to the user's preferred kind of activities. When the OWA-based system is used and the semantic criterion does not use the veto power (*OWA-Avoid SEM* column), there are many shopping activities in the positions 3-12

Table 5.9: PROFILE 1: SWR and OWA rankings using veto in all criteria

R	SWR-ALL	NFS	OWA-ALL	NFS
1	Platja de l'Arrabassada {NormalBeaches, BeachPicnic, PicnicAreas, FamilyBeaches, RefreshmentStall }	57,78	Platja Cala Romana {NormalBeaches, BeachPicnic}	59,744
2	Platja Cala Romana {NormalBeaches, BeachPicnic}	56,702	Platja de l'Arrabassada {NormalBeaches, BeachPicnic, PicnicAreas, FamilyBeaches, RefreshmentStall }	55,134
3	Les Gavarres {ShoppingArea, FoodAreas, Pizzeria, InternationalCousine }	51,956	Xiringuito Xaloc {RefreshmentStall, Pubs, TastingCousing, TapasCousine }	45,196
4	Xiringuito Xaloc {RefreshmentStall, Pubs, TastingCousing, TapasCousine }	48,592	Les Gavarres {ShoppingArea, FoodAreas, Pizzeria, InternationalCousine }	45,075
5	Casino (Rambla Nova) {GameRoom, Bars, Discos }	40,466	Carrer Major {LocalProducts, LocalMarket}	39,228
6	Carnaval de Tarragona {TraditionalCelebrations, DanceFestivals, MusicFestivals, BigGroupsAtmosphere }	37,652	El Corte Ingles {ShoppingCenter, ShoppingArea }	39,1
7	Platja Sabinosa {NudismBeaches, CoastalRoutes }	37,332	Mercadet de Tarragona {LocalProducts, ArtsAndCraftsEvents }	35,41
8	Primera setmana {TraditionalCelebrations, PopularCelebrations }	32,776	Platja Sabinosa {NudismBeaches, CoastalRoutes }	34,716
9	Mercadet de Tarragona {LocalProducts, ArtsAndCraftsEvents }	32,415	Mercat Torreforta { LocalMarket, LocalProducts }	32,204
10	Festa oli DOP Siurana {TraditionalCousine, WineFestivals, WineFairs, PopularCelebrations, OilRoutes }	31,134	Platja de la Pineda {NormalBeaches, UrbanBeaches}	31,574
11	Santa Tecla {TraditionalCelebrations, LocalProducts, WineAtmosphere, PopularCelebrations, MusicFestivals, ArtsAndCraftsEvents, BigGroupsAtmosphere}	30,629	Via T {ShoppingArea, LocalProducts, FoodAreas }	30,969
12	Casino (Rambla Vella) {GameRoom, BallRoom, Bars}	29,373	Mercat Central Taragona {LocalMarket, LocalProducts, OutstandingProduct, UniqueBuilding}	30,944
13	Platja de la Pineda {NormalBeaches, UrbanBeaches}	28,707	Parc Central {ShoppingCenter, ShoppingArea }	29,09
14	Divendres Sant {TraditionalCelebrations, PopularCelebrations, Cathedral }	27,08	Primera setmana {TraditionalCelebrations, PopularCelebrations }	28,989
15	Carrer Major {LocalProducts, LocalMarket }	25,44	Festa oli DOP Siurana {TraditionalCousine, WineFestivals, WineFairs, PopularCelebrations, OilRoutes }	28,965
16	Mercat Central Taragona {LocalMarket, LocalProducts, OutstandingProduct, UniqueBuilding}	25,217	Mercadet de Constant {LocalMarket, LocalProducts}	25,634
17	El Corte Ingles {ShoppingCenter, ShoppingArea }	24,209	Pedrera Romana Medol {HistoryMuseums, HumanHeritage, Roman, ArcheologyMuseums, Ruins, CultureRoutes }	24,639
18	Via T {ShoppingArea, LocalProducts, FoodAreas }	21,226	Xiringuito Sol Solet {RefreshmentStall, Bars, TapasCousine}	23,931
19	Mercat Torreforta {LocalMarket, LocalProducts }	19,429	Museu dels Fars {SeaportMuseums}	23,449
20	Xiringuito Sol Solet {RefreshmentStall,Bars,TapasCousine}	18,008	Museu Art Modern {ArtMuseums,PaintingMuseums }	23,214

5.3. A recommender system for tourists visiting Costa Daurada & Terres de l'Ebre 117

and cultural activities in the 15-20 interval, which are not very adequate for this user. However, in the*SWR-Avoid SEM* case, although there are still shopping activities, we can find several events in the positions 13-17, which fit better the user's interests.

When only the semantic criterion may veto (but the 3 numerical ones can not use their veto power -*Avoid NUM* columns), the price or the distance are not able to veto the semantic tag interest scores. In this case, we can see that some expensive options appear in the list of SWR method, like casinos (positions 7, 9 and 14), as well as others that are far away from Tarragona city, like visiting cellars (positions 17 and 19). The SWR ranking displays several gastronomy-related activities in the positions 4-9, and several events in the 10-15 interval, fitting nicely with the user's interests. On the contrary, the OWA ranking still keeps many shopping activities (positions 11-15). Thus, in this case the best ranking seems to be the one generated by SWR when only the semantic criterion has veto power.

Table 5.10: PROFILE 1: SWR and OWA ranking without veto on some criteria

R	SWR-Avoid SEM	OWA-Avoid SEM	SWR-Avoid NUM	OWA-Avoid NUM
1	Platja de l'Arrabassada	Platja de l'Arrabassada	Platja de l'Arrabassada	Platja Cala Romana
2	Platja Cala Romana	Platja Cala Romana	Platja Cala Romana	Platja de la Pineda
3	Les Gavarres	Les Gavarres	Platja de la Pineda	Platja de l'Arrabassada
4	Mercadet de Tarragona	El Corte Ingles	Xiringuito Xaloc	Xiringuito Sol Solet
5	Carrer Major	Carrer Major	Les Gavarres	Xiringuito Xaloc
6	El Corte Ingles	Mercat Central Taragona	Xiringuito Sol Solet	Xiringuito Casablanca
7	Mercat Central Taragona	Mercadet de Tarragona	Casino (Rambla Vella)	Les Gavarres
8	Platja Sabinosa	Mercat Torreforta	Xiringuito Casablanca	Primera setmana
9	Mercat Torreforta	Platja Sabinosa	Casino (Rambla Nova)	Festa oli DOP Siurana
10	Via T	Via T	Carnaval de Tarragona	Platja Sabinosa
11	Parc Central	Parc Central	Primera setmana	El Corte Ingles
12	Mercadet de Constant	Mercadet de Constant	Festa oli DOP Siurana	Carrer Major
13	Festa oli DOP Siurana	Primera setmana	Santa Tecla	Via T
14	Primera setmana	Festa oli DOP Siurana	Casino Sant Salvador	Parc Central
15	Carnaval de Tarragona	Pedrera Romana Medol	Divendres Sant	Mercadet de Tarragona
16	Divendres Sant	Museu dels Fars	Platja Sabinosa	Casino (Rambla Nova)
17	Santa Tecla	Museu Art Modern	Mas dels Frares	Divendres Sant
18	Xiringuito Xaloc	Museu del Port	La Pineda. Discoteca	Mercat Torreforta
19	Museu del Port	Museu Necropolis	Celler La Boella	Casino (Rambla Vella)
20	Pedrera Romana Medol	Muralles Tarragona	Mercadet de Tarragona	Carnaval de Tarragona

Results for PROFILE 2

The second profile is for a group of tourists that are focused on relax, Nature and, to a lesser extent, in Leisure activities. In second term, they like sports, gastronomy and popular events. They do not like shopping, going to the beach or making cultural activities.

The top 6 positions in the two rankings for profile 2 (Table 5.11) are for cheap activities that mix leisure and gastronomy, which are two of the preferred items (with 52% and 20% respectively). We can see Carrer Major (Main Street in Tarragona with shops and local products) and several markets (like Mercat Central or Mercadet de Tarragona). These activities have probably been selected due to the fact that the tags LocalProducts and LocalMarket are linked to gastronomy, while ArtsAndCraftsEvents corresponds to Leisure activities.

The OWA ranking does not provide any relax (90%) or Sports (23%) activities, which may be found in the SWR ranking (19-20 Relax, 16 Sports). Both rankings display traditional events, although they are more highly valued in the SWR ranking (7, 12-18) than in the OWA one 13-14, 18-20). The OWA ranking shows a cultural activity in position 7 (highly disliked by the user, 0% interest), which does not appear in the SWR ranking. The OWA ranking only shows 1 Nature event (85% interest) in position 12, whereas the SWR list shows this kind of events in positions 8 and 11. Thus, in general, it may be seen that the SWR ranking seems to fit better much better the user's interests than the previous OWA one (Table 5.12).

If we analyze the results in the case in which the semantic criterion does not have veto power, we find similar rankings for OWA and SWR, with the top 10 positions mainly covered by leisure activities (52% interest), and the 12-20 interval by popular events (18% events). As the Relax activities (90%) are quite expensive, they don't appear in these lists. There are 3 beaches and 1-2 cultural activities, despite the low interest of the user (5% and 0%); however, there are not specific Nature-related activities (85%), probably because they are far from the city centre.

The results are quite different when the numerical criteria like price or distance do not have veto power. In this case there are Nature activities in the top 2 positions of the SWR ranking (positions 1 and 10 in OWA) and also Relax activities (3, 11 and 15 in SWR, 3, 9 and 13 in OWA). A Sport

5.3. A recommender system for tourists visiting Costa Daurada & Terres de l'Ebre 119

Table 5.11: PROFILE 2: SWR and OWA rankings using veto in all criteria

R	SWR-ALL	NFS	OWA-ALL	NFS
1	Carrer Major LocalProducts, LocalMarket	38,026	Carrer Major LocalProducts, LocalMarket	49,689
2	Mercat Central Taragona LocalMarket, LocalProducts, OutstandingProduct, UniqueBuilding	37,054	Mercat Central Taragona LocalMarket, LocalProducts, OutstandingProduct, UniqueBuilding	47,444
3	Mercadet de Tarragona LocalProducts, ArtsAndCraftsEvents	33,22	Mercat Torreforta LocalMarket, LocalProducts	44,097
4	Mercat Torreforta LocalMarket, LocalProducts	33,106	Mercadet de Tarragona LocalProducts, ArtsAndCraftsEvents	42,178
5	Via T ShoppingArea, LocalProducts, FoodAreas	29,468	Mercadet de Constant LocalMarket, LocalProducts	38,125
6	Mercadet de Constant LocalMarket, LocalProducts	27,779	El Corte Ingles ShoppingCenter, ShoppingArea	33,5
7	Carnaval de Tarragona TraditionalCelebrations, DanceFestivals, MusicFestivals, BigGroupsAtmosphere	27,006	Museu dels Fars SeaportMuseums	31,974
8	Sequia Major NaturalSpaces, RuralViewPoints, Forests, PicnicAreas, FaunaReserves, GorgeWalking	24,945	Via T ShoppingArea, LocalProducts, FoodAreas	30,644
9	Les Gavarres ShoppingArea, FoodAreas, Pizzeria, InternationalCousine	24,869	Les Gavarres ShoppingArea, FoodAreas, Pizzeria, InternationalCousine	25,445
10	Xiringuito Xaloc RefreshmentStall, Pubs, TastingCousing, TapasCousine	23,068	Parc Central ShoppingCenter, ShoppingArea	23,29
11	Cala de la Roca Plana Coves, NaturalSpaces, NudismBeaches, InlandWatersRoutes, CoastalRoutes, ViaFerrata	21,483	Platja de l'Arrabassada NormalBeaches, BeachPicnic, PicnicAreas, FamilyBeaches, RefreshmentStall	22,89
12	Primera setmana TraditionalCelebrations, PopularCelebrations	20,867	Sequia Major NaturalSpaces, RuralViewPoints, Forests, PicnicAreas, FaunaReserves, GorgeWalking	22,218
13	Santa Tecla TraditionalCelebrations, LocalProducts, WineAtmosphere, PopularCelebrations, MusicFestivals, ArtsAndCraftsEvents, BigGroupsAtmosphere	19,842	Primera setmana TraditionalCelebrations, PopularCelebrations	21,839
14	Festa oli DOP Siurana TraditionalCousine, WineFestivals, WineFairs, PopularCelebrations, OilRoutes	18,932	Festa oli DOP Siurana TraditionalCousine, WineFestivals, WineFairs, PopularCelebrations, OilRoutes	20,306
15	Celler Mas dels Frares WineTasting	18,53	Celler Mas dels Frares WineTasting	20,189
16	Rocodrom CE ClimbingWall, RockClimbing	18,013	Platja Sabinosa NudismBeaches, CoastalRoutes	18,7
17	Sant Mag TraditionalCelebrations, PopularCelebrations, Chapel, RuralAtmosphere	16,97	Platja Cala Romana NormalBeaches, BeachPicnic	16,768
18	Divendres Sant TraditionalCelebrations, PopularCelebrations, Cathedral	16,898	Carnaval de Tarragona TraditionalCelebrations, DanceFestivals, MusicFestivals, BigGroupsAtmosphere	16,674
19	Aquum Spa & Club HydroFacilities, SpaResorts, MineralWaterResorts	13,833	Santa Tecla TraditionalCelebrations, LocalProducts, WineAtmosphere, PopularCelebrations, MusicFestivals, ArtsAndCraftsEvents, BigGroupsAtmosphere	16,569
20	Venus Wellness Center HealthResorts, HydroFacilities, SpaResorts	13,028	Sant Magi TraditionalCelebrations, PopularCelebrations, Chapel, RuralAtmosphere	14,684

Table 5.12: PROFILE 2: SWR and OWA ranking without veto on some criteria

R	SWR-Avoid SEM	OWA-Avoid SEM	SWR-Avoid NUM	OWA-Avoid NUM
1	Carrer Major	Carrer Major	Sequia Major	Sequia Major
2	Mercat Central Taragona	Mercat Central Taragona	Cala de la Roca Plana	Venus Wellness Center
3	Mercadet de Tarragona	Mercat Torreforta	Venus Wellness Center	Celler Mas dels Frares
4	Mercat Torreforta	Mercadet de Tarragona	Celler Mas dels Frares	Carrer Major
5	Mercadet de Constant	Mercadet de Constant	Carrer Major	Mercat Central Taragona
6	Via T	El Corte Ingles	Carnaval de Tarragona	Mercat Torreforta
7	Les Gavarres	Via T	Via T	Museu dels Fars
8	El Corte Ingles	Museu dels Fars	Mercat Central Taragona	Via T
9	Platja Sabinosa	Les Gavarres	Mercat Torreforta	Aquum Spa & Club
10	Parc Central	Parc Central	Dalmau Hnos y cia	Cala de la Roca Plana
11	Platja de l'Arrabassada	Platja de l'Arrabassada	Aquum Spa & Club	Mercadet de Tarragona
12	Festa oli DOP Siurana	Primera setmana	Xiringuito Xaloc	Mercadet de Constant
13	Primera setmana	Festa oli DOP Siurana	Mercadet de Tarragona	Spalas (Gran Palas Hotel)
14	Carnaval de Tarragona	Platja Sabinosa	Primera setmana	El Corte Ingles
15	Santa Tecla	Platja Cala Romana	Spalas (Gran Palas Hotel)	Primera setmana
16	Museu Biblic	Carnaval de Tarragona	Rocodrom CE	Festa oli DOP Siurana
17	Divendres Sant	Santa Tecla	Festa oli DOP Siurana	Parc Central
18	Sant Mag	Pedrera Romana Medol	Santa Tecla	Carnaval de Tarragona
19	Platja Cala Romana	Sant Mag	Mercadet de Constant	Santa Tecla
20	Xiringuito Xaloc	Divendres Sant	Celler La Boella	Les Gavarres

(23% interest) also appears in the SWR ranquing (16), and a Cultural activity (0%) still appears in the OWA ordering. Thus, in this case the orderings that avoid the numerical veto, and especially the SWR-based one, seem to provide a more appropriate set of recommendations.

Results for PROFILE 3

The third profile corresponds to a group of youths interested mainly in culture (76%), gastronomy (55%) and events (45%). They do not like sports, going to the beach or Nature-related activities.

In this case, the recommender system displays similar top alternatives in both methods, showing museums and other cultural points of interest in the top 13 positions (Table 5.13). The city of Tarragona has a lot of cultural places, mainly related to its origin as the capital of a province of the Roman empire in the Iberian peninsula. From position 14 to 20 both methods show some markets where tourists can buy local products and some shopping areas with lots of restaurants (e.g. Carrer Major, Via T). It is worth noting that the SWR ranking is able to retrieve the *"Tarraco Viva"'* event that fits with two of the interests of the user, while it is not recommended when using OWA.

5.3. A recommender system for tourists visiting Costa Daurada & Terres de l'Ebre

Table 5.13: PROFILE 3: SWR and OWA rankings using veto in all criteria

R	SWR-ALL	NFS	OWA-ALL	NFS
1	Pedrera Romana Medol HistoryMuseums, HumanHeritage, Roman, ArcheologyMuseums, Ruins, CultureRoutes	53,299	Museu dels Fars SeaportMuseums	52,905
2	Museu dels Fars SeaportMuseums	52,33	Museu Art Modern ArtMuseums, PaintingMuseums	52,797
3	Museu Art Modern ArtMuseums, PaintingMuseums	52,222	3 Museu Casa Castellarnau HistoryMuseums, Roman, InterpretationCenter	52,7
4	Muralles HistoryMuseums, HumanHeritage, Roman, Walls	52,06	Museu Biblic SacredArtMuseums, Roman SacredArtMuseums, Roman	52,635
5	Museu Casa Castellarnau HistoryMuseums, Roman, InterpretationCenter	51,975	Museu Necropolis ArcheologyMuseums, HumanHeritage, Roman	52,624
6	Museu Biblic SacredArtMuseums, Roman	51,96	Amfiteatre Roma Amphitheater, HistoryMuseums, HumanHeritage, Roman, Romanesque, Ruins	52,585
7	Amfiteatre Roma Amphitheater, HistoryMuseums, HumanHeritage, Roman, Romanesque, Ruins	51,937	Museu del Port HistoryMuseums, HistoricBuilding, SeaportMuseums	52,508
8	Museu Necropolis ArcheologyMuseums, HumanHeritage, Roman	51,932	Museu Diocesa Tarragona ArtMuseums, Roman	52,5
9	Museu del Port HistoryMuseums, HistoricBuilding, SeaportMuseums	51,883	Pedrera Romana Medol HistoryMuseums, HumanHeritage, Roman, ArcheologyMuseums, Ruins, CultureRoutes	52,353
10	Museu Diocesa Tarragona ArtMuseums, Roman	51,725	Pretori i Forum Provincial Gothic, HistoryMuseums, HumanHeritage, Palace, Roman	50,365
11	Pretori i Forum Provincial Gothic, HistoryMuseums, HumanHeritage, Palace, Roman	51,604	CIRC ROM Circus, HistoryMuseums, HumanHeritage, Roman	47,274
12	Circ Roma Circus, HistoryMuseums, HumanHeritage, Roman	47,049	Mercadet de Tarragona LocalProducts, ArtsAndCraftsEvents	42,972
13	Vila romana Centcelles ArcheologyMuseums, HumanHeritage, Roman	41,497	Vila romana Centcelles ArcheologyMuseums, HumanHeritage, Roman	41,339
14	Mercadet de Tarragona LocalProducts, ArtsAndCraftsEvents	38,699	Conjunt roma de Tarragona Amphitheater, Aqueduct, Circus, HumanHeritage, Roman, Tower, Walls	38,339
15	Conjunt roma de Tarragona Amphitheater, Aqueduct, Circus, HumanHeritage, Roman, Tower, Walls	37,565	Carrer Major LocalProducts, LocalMarket	35,715
16	Mercat Central Taragona LocalMarket, LocalProducts, OutstandingProduct, UniqueBuilding	32,681	Mercat Torreforta LocalMarket, LocalProducts	30,914
17	Carrer Major LocalProducts, LocalMarket	31,432	El Corte Ingles ShoppingCenter, ShoppingArea	30
18	Tarraco Viva TraditionalCelebrations, Roman, Medieval, GastronomyFestivals	30,702	Via T ShoppingArea, LocalProducts, FoodAreas	27,158
19	Via T ShoppingArea, LocalProducts, FoodAreas	28,453	Mercat Central Taragona LocalMarket, LocalProducts, OutstandingProduct, UniqueBuilding	26,924
20	Mercat Torreforta LocalMarket, LocalProducts	27,754	Mercadet de Constant LocalMarket, LocalProducts	26,881

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

122   **Chapter 5.   Tools and applications**

In Table 5.14 it is possible to notice some differences depending on which criteria use the veto power. When the touristic tags (i.e. the semantic criterion) is not using the veto, the first positions are for cheap activities like visiting markets or shopping streets (notice that we do not consider the cost of buying things), while the museums that have a ticket price start from position 6. However, when the price cannot veto, then the best alternatives (positions 1-14) are the museums, despite its price. This is the expected behaviour of the veto threshold, which is working well with both the SWR and OWA methods. The *"Tarraco Viva"* event also appears in this case (SWR-16, OWA-19). Therefore, in this case it could be argued that the rankings without numeric veto are the best ones.

Table 5.14: PROFILE 3: SWR and OWA ranking without veto on some criteria

R	SWR-Avoid SEM	OWA-Avoid SEM	SWR-Avoid NUM	OWA-Avoid NUM
1	Mercadet de Tarragona	Carrer Major	Pedrera Romana Medol	Museu dels Fars
2	Carrer Major	Mercadet de Tarragona	Museu dels Fars	Museu Art Modern
3	Pedrera Romana Medol	El Corte Ingles	Museu Art Modern	Museu Casa Castellarnau
4	Mercat Central Taragona	Pedrera Romana Medol	Muralles	Muralles
5	El Corte Ingles	Mercat Central Taragona	Museu Casa Castellarnau	Museu Biblic
6	Museu dels Fars	Museu dels Fars	Museu Biblic	Museu Necropolis
7	Museu Art Modern	Mercat Torreforta	Amfiteatre Roma	Amfiteatre Roma
8	Muralles	Museu Art Modern	Museu Necropolis	Museu del Port
9	Museu Necropolis	Muralles	Museu del Port	Museu Diocesa
10	Museu del Port	Museu del Port	Museu Diocesa Tarragona	Pedrera Romana Medol
11	Museu Casa Castellarnau	Museu Necropolis	Pretori i Forum Provincial	Pretori i Forum Provincial
12	Museu Biblic	Museu Casa Castellarnau	Circ Roma	Circ Roma
13	Amfiteatre Roma	Museu Biblic	Vila romana Centcelles	Conjunt roma de Tarragona
14	Via T	Amfiteatre Roma	Conjunt roma de Tarragona	Vila romana Centcelles
15	Museu Diocesa	Museu Diocesa Tarragona	Mercadet de Tarragona	Mercadet de Tarragona
16	Pretori i Forum Provincial	Via T	Tarraco Viva	Torre Ermita La Pineda
17	Mercat Torreforta	Pretori i Forum Provincial	Torr e Ermita La Pineda	Carrer Major
18	Circ Roma	Circ Roma	Via T	Via T
19	Les Gavarres	Mercadet de Constant	Mercat Central Taragona	Tarraco Viva
20	Mercadet de Constant	Parc Central	Carrer Major	Mercat Torreforta

Results for PROFILE 4

The last profile considered in this case study corresponds to a group
of visitors interested mainly on going to the beach (80%), shopping (54%)
and sports (54%). They have a more diverse profile, as they also like relax
(33%), Nature (38%) and Events (30%). They want to avoid cultural (6%)
and Leisure (7%) activities.

In this situation (Table 5.15), the first two options of the SWR ranking
correspond to beach activities. They are mixed with shopping centers
in the rest of the list until position 10. Then, some sport activities are
recommended, like scuba diving, rock climbing, kayaking or sailing. We
can also find the activity "Cala de la Roca Plana" in position 18, which
includes both natural spaces with beach as well as some sport (via ferrata).

In the OWA recommendations, on the other hand, the top position is
for a gastronomy and shopping activity, which does not fit so nicely the
user's interests. The activity "Cala de la Roca Plana" appears in position
20, later than with SWR. Sport activities are similar to the SWR ranking,
with some rank reversals in the order.

Table 5.16 shows the results when some of the criteria can not apply
veto. If the touristic tags semantic criterion can't veto, the top 12 positions
of the OWA and SWR lists are filled with shopping places and beaches.
The OWA ranking shows 4 cultural events at the bottom, which would
not be appropriate for this user. The SWR is a little bit better because
it has more Events in the positions 13-20, rather than cultural activities.
Unfortunately, none of the two rankings includes any sports activity.

When the numerical attributes can not veto the outranking relation,
the becahes occupy most of the top positions, especially in the SWR case.
The SWR ranking now features 7 sports activities (5 in OWA). The only
Relax activity appears in position 16 of the OWA list.

These results seem to confirm that SWR is providing more accurate
results according to the set of tags of each activity and the corresponding
user's preferences, especially when the numerical criteria can't veto the
final result.

Table 5.15: PROFILE 4: SWR and OWA rankings using veto in all criteria

R	SWR-ALL	NFS	OWA-ALL	NFS
1	Platja Cala Romana NormalBeaches, BeachPicnic	57,119	Carrer Major LocalProducts, LocalMarket	61,843
2	Platja de l'Arrabassada NormalBeaches, BeachPicnic, Pic- nicAreas, FamilyBeaches, Refresh- mentStall	55,176	Platja Cala Romana NormalBeaches, BeachPicnic	58,694
3	El Corte Ingles ShoppingCenter, ShoppingArea	49,475	El Corte Ingles ShoppingCenter, ShoppingArea	56,975
4	Parc Central ShoppingCenter, ShoppingArea	45,928	Parc Central ShoppingCenter, ShoppingArea	53,427
5	Platja Sabinosa NudismBeaches, CoastalRoutes	45,731	Platja de l'Arrabassada NormalBeaches, BeachPicnic, Pic- nicAreas, FamilyBeaches, Refresh- mentStall	50,907
6	Carrer Major LocalProducts, LocalMarket	43,664	Platja Sabinosa NudismBeaches, CoastalRoutes	46,206
7	Mercat Torreforta LocalMarket, LocalProducts	38,074	Mercat Torreforta LocalMarket, LocalProducts	39,018
8	Mercadet de Constant LocalMarket, LocalProducts	31,589	Casino (Rambla Nova) GameRoom, Bars, Discos	32,585
9	Casino (Rambla Nova) GameRoom, Bars, Discos	29,768	Mercadet de Constant LocalMarket, LocalProducts	31,916
10	Platja de la Pineda NormalBeaches, UrbanBeaches	29,374	Platja de la Pineda NormalBeaches, UrbanBeaches	31,449
11	Submarinisme ScubaDiving, Snorkelling	18,103	Mercat Central Taragona LocalMarket, LocalProducts, Out- standingProduct, UniqueBuilding	20,494
12	Mercat Central Taragona LocalProducts, ArtsAndCraftsEv- ents	17,818	Submarinisme ScubaDiving, Snorkelling	19,478
13	Busseig a Escullera Snorkelling, ScubaDiving	16,721	Mercadet de Tarragona LocalProducts, ArtsAndCraftsEv- ents	18,868
14	Rocodrom CE ClimbingWall, RockClimbing	16,038	Via T ShoppingArea, LocalProducts, FoodAreas	18,244
15	Casino (Rambla Vella) GameRoom, BallRoom, Bars	15,863	Rocodrom CE ClimbingWall, RockClimbing	18,196
16	Via T ShoppingArea, LocalProducts, FoodAreas	14,95	Casino (Rambla Vella) GameRoom, BallRoom, Bars	17,584
17	Caiac. Platja Llarga Kayaking, Canoeing	14,776	Caiac. Platja Llarga Kayaking, Canoeing	16,851
18	Cala de la Roca Plana Coves, NaturalSpaces, Nudism- Beaches, InlandWatersRoutes, CoastalRoutes, ViaFerrata	14,232	Busseig a Escullera Snorkelling, ScubaDiving	16,571
19	Mercadet de Tarragona LocalProducts, ArtsAndCraftsEv- ents	13,17	Sant Magi TraditionalCelebrations, Popu- larCelebrations, Chapel, RuralAt- mosphere	13,659
20	Vela Litoral Tarragona Boating, Sail, Windsurfing, Inland- WatersRoutes	10,555	Cala de la Roca Plana Coves, NaturalSpaces, Nudism- Beaches, InlandWatersRoutes, CoastalRoutes, ViaFerrata	11,397
20	Vela Litoral fins Torredemb Boating, Sail, Windsurfing, Inland- WatersRoutes			

Table 5.16: PROFILE 4: SWR and OWA ranking without veto on some criteria

R	SWR-Avoid SEM	OWA-Avoid SEM	SWR-Avoid NUM	OWA-Avoid NUM
1	El Corte Ingles	Carrer Major	Platja Cala Romana	Carrer Major
2	Carrer Major	El Corte Ingles	Platja de l'Arrabassada	Platja Cala Romana
3	Platja de l'Arrabassada	Platja de l'Arrabassada	Platja de la Pineda	Platja de la Pineda
4	Platja Cala Romana	Platja Cala Romana	El Corte Ingles	El Corte Ingles
5	Mercat Torreforta	Mercat Torreforta	Platja Sabinosa	Parc Central
6	Platja Sabinosa	Platja Sabinosa	Parc Central	Platja de l'Arrabassada
7	Parc Central	Parc Central	Busseig a Escullera	Busseig a Escullera
8	Mercadet de Constant	Mercadet de Constant	Cala de la Roca Plana	Platja Sabinosa
9	Mercat Central Taragona	Mercat Central Taragona	Carrer Major	Casino (Rambla Nova)
10	Via T	Mercadet de Tarragona	Mercat Torreforta	Submarinisme
11	Mercadet de Tarragona	Via T	Casino (Rambla Nova)	Cala de la Roca Plana
12	Les Gavarres	Les Gavarres	Submarinisme	Casino (Rambla Vella)
13	Sant Magi	Sant Magi	Casino (Rambla Vella)	Caiac. Platja Llarga
14	Primera setmana	Pedrera Romana Medol	Mercadet de Constant	Mercat Torreforta
15	Festa oli DOP Siurana	Primera setmana	Caiac. Platja Llarga	Rocodrom CE
16	Carnaval de Tarragona	Casino (Rambla Nova)	Rocodrom CE	Venus Wellness Center
17	Pedrera Romana Medol	Museu dels Fars	Vela Litoral Tarragona	Mercadet de Constant
18	Casino (Rambla Nova)	Museu Art Modern	Rem al Litoral Tarragona	Via T
19	Museu dels Fars	Muralles Tarragona	Mercadet de El Morell	Sant Magi
20	Divendres Sant	Museu del Port	Creuer al Port TGN	Creuer al Port TGN

Overall comparison

Although the previous tables only display the 20 top alternatives of the rankings, we have actually compared the ranking positions of the 70 activities of this case study in the different user profiles. Table 5.17 shows the Pearson correlation coefficients obtained in each case.

When semantic criteria do not use discordance (i.e. veto) the rankings obtained are quite similar, just with small rank reversals. On the contrary, when the numerical criteria do not use the veto (*Avoid NUM*), the correlation coefficient is smaller than in the other cases (except in profile 1 which is quite close to the normal case, in which all criteria veto). As mentioned before, this was the expected behavior of the test, because when the result is based to a higher extent in the semantic criterion concordance and discordance indices, SWR and OWA may give significantly different results in some user profiles (like profile 2). OWA is losing information during the aggregation stage, while SWR is managing the tag interest scores in a more appropriate way.

Table 5.17: Correlation between method SWR and OWA

	Profile 1	Profile 2	Profile 3	Profile 4
All	0,896	0,666	0,985	0,922
Avoid SEM	0,982	0,987	0,996	0,989
Avoid NUM	0,908	0,592	0,979	0,899

5.4   Summary

In this chapter two different application fields have been used to show
the powerand generality of the methods proposed in this doctoral thesis.
The new SWR-based management of semantic criteria has been introduced
in a software tool developed in my research group ITAKA. The authorship
of this tool has been registered at the Benelux Office for Intellectual
Property (with an i-DEPOT registration) and it is being disseminated
and commercialized by Universitat Rovira i Virgili through the INNOGET
platform (www.innoget.com).

The first case study has shown a different modeling of the semantic
data by using a more complex user profile made upon utility functions
associated to the concepts of the ontology. In that way, the tag interest
scores are not fixed but depend on the alternative to which they are linked.
In this case study we also modeled the scores as risk factors, having thus
to minimize the score instead of maximizing it. It has been proven that
the proposed method is able to deal with these particularities.

Dealing with an environmental problem has been very appealing because
MCDA methods are now widely applied to this kind of problems. Our work,
then, shows that the ELECTRE-SEM method can also be useful in this
area. The experts that collaborated in this case study were very satisfied
with the obtained results and with the possibility of managing semantic
data in this kind of complex decision environments. The definition of
ad-hoc ontologies has also been a good contribution as they can be reused
in other applications dealing with pollutants.

5.4. Summary 127

In the second case study we have relied on a previous recommender system in the field of Tourism management. This has given us the opportunity to make a comparison between our method, ELECTRE-SEM, and another approach based on aggregation operators. The comparison consisted in 4 profiles of tourists and we observed that the aggregation of interest scores into a single overall score may produce sometimes a loss of information. Our method, instead, makes more appropriate recommendations according to the preferences of each kind of user.

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

6

Conclusions and future work

In this chapter, we conclude our work with a summary of the contributions and some suggestions for future work.

6.1 Conclusions

Decision making is a very hard task, which often requires the analysis of hundreds or thousands of potential alternatives defined on multiple conflicting attributes. Methods based on decision rules, utility functions or outranking methods have been proposed in the *Multi Criteria Decision Aid (MCDA)* field. Most of these methods deal only with numerical and categorical attributes, so a current research challenge is the incorporation of new kinds of information, like semantic attributes. The solution to this problem requires the definition of a way of representing the user preferences on the values of these attributes (which are concepts in a domain ontology), and the use of this preferential information in a sound MCDA methodology (in our case, the well-known ELECTRE family of outranking methods).

This dissertation has proposed the use of ontologies to represent the preferences of the user by means of a numerical indicator, the *Tag Interest Score (TIS)*, associated to each leaf of the ontology. One of the contributions of this dissertation has been the definition of a novel preference learning mechanism, which may infer the preference on a concept by analyzing the user's interest on semantically similar concepts. This algorithm is based on a WOWA-based aggregation, which includes the novelty of considering two

sets of weights to guide the aggregation. One of them captures the semantic similarity of the concepts, which is calculated from the structure of a domain ontology associated to the semantic criterion. The other one is employed to avoid undesired compensatory effects, by defining a conjunctive or disjunctive aggregation policy. This mechanism was described in chapter 2, where its performance on different types of user profiles was studied. It was seen that the best aggregation policy (optimistic or pessimistic) depends on the characteristics of the user. The method was also compared with other algorithms from the state of the art, showing a similar performance.

One of the basic contributions of the thesis is the enhancement of the classic ELECTRE method so that it can deal with multi-valued semantic attributes. The basic idea has been the definition of new concordance and discordance functions on this kind of criteria. These functions are based on the concept of *Semantic Win Rate*, which is a new way of comparing the performance of two alternatives on a semantic criterion. The resulting system, ELECTRE-SEM, described in chapter 4, keeps the general spirit of ELECTRE (with preference, indifference and veto thresholds) and it allows to work with numerical, categorical and multi-valued semantic criteria.

The work in the dissertation has not only been theoretical. The new contributions have been implemented in a software package, ELECTRE-SEM, which can work with linear or hierarchical sets of criteria. The system has been applied to two real-world domains which are considered of special relevance in URV: Tourism and Environment. In the first case, we have compared the performance of the new system with the one of a previous recommender system of tourist activities developed in conjunction with the Science and Technology Park for Tourism and Leisure. Our qualitative comparative study shows that the new system produces recommendations that fit better with the user's preferences. In the second case, in my stay at Oxford Brookes University I applied the system to assess different kinds of power generation plants, considering economic and environmental criteria, finding that the renewable procedures have a better performance. In particular, for the case of UK, Wind and Biomass power plants are the most environmentally sustainable and the cheapest to maintain.

6.2   Future work

The work done in this thesis opens an interesting and promising research line that enables the use of semantic data for decision support. I propose the following topics of study to continue the work stated in this dissertation:

- The Semantic Win Rate measure, together with the definitions of concordance and discordance indices for semantic criteria, could be easily adapted to other outranking-based MCDA methods, like PROMETHEE, presented in chapter 3. Moreover, the idea of ontology-based user profiles (with Tag Interest Scores) could be also included in other kind of MCDA methods based on comparisons of alternatives, like TOPSIS.

- Obtaining good results with ELECTRE-SEM requires the adjustment of some preference parameters (i.e. indifference, preference and veto thresholds, as well as the minimum semantic win rate). As acknowledged in the literature of MCDA, decisions makers sometimes have difficulties in providing directly all these values. There exist some methods that obtain the values of the parameters through indirect elicitation, like the Robust Ordinal Regression technique (ROR) where the user provides information by comparing some reference alternatives to obtain a set of compatible preference parameters. It would be worth to study the extension of current ROR techniques for the new parameters of the semantic criteria defined in this thesis.

- In this thesis we have made a qualitative comparison of the results of ELECTRE-SEM with the previous version of a tourism recommender system (SigTur). In the future, with the collaboration of the Science and Technology Park for Tourism and Leisure, we would like to plan a testing of the new techniques with real visitors of the area of Tarragona. To make this possible, the new algorithms developed in this thesis must be integrated into the SigTur system, substituting the previous ones. In addition to observing the tourists' behaviour in the system, it would be necessary to define an appropriate way of evaluating the user's satisfaction by means of a questionnaire in order to know the perception of the tourist's on the lists of recommended activities.

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

Bibliography

[1] F. H. Abanda, J. H. M. Tah, and D. Duce. PV-TONS: A photovoltaic technology ontology system for the design of PV-systems. *Engineering Applications of Artificial Intelligence*, 26(4):1399–1412, 2013.

[2] G. Adomavicius and Y. Kwon. New recommendation techniques for multicriteria rating systems. *IEEE Intelligent Systems*, 22(3):48–55, 2007.

[3] N. H. Afgan and M. G. Carvalho. Multi-criteria assessment of new and renewable energy power plants. *Energy*, 27(8):739–755, 2002.

[4] A. Aldea, R. Bañares-Alcántara, and S. Skrzypczak. Managing information to support the decision making process. *Journal of Information & Knowledge Management*, 11(03):1250016, 2012.

[5] S. S. Anand, P. Kearney, and M. Shapcott. Generating semantically enriched user profiles for web personalization. *ACM Transactions on Internet Technology*, 7(4), 2007.

[6] M. Archana, A. K, S. Apoorva, and A. J. A survey on existing tourism ontologies. *International Journal of Research in Engineering and Technology*, 4(14):20–23, 2015.

[7] M. Bhatt, W. Rahayu, S. P. Soni, and C. Wouters. Ontology driven semantic profiling and retrieval in medical information systems. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(4):317–331, 2009.

[8] Y. Blanco-Fernández, M. López-Nores, A. Gil-Solla, M. Ramos-Cabrer, and J. J. Pazos-Arias. Exploring synergies between content-based filtering and spreading activation techniques in knowledge-based recommender systems. *Information Sciences*, 181(21):4823–4846, 2011.

[9] J. Borràs. *Semantic Recommender Systems Provision of personalised information about tourist activities*. PhD thesis, Universitat Rovira i Virgili, 2015.

[10] J. Borràs, A. Moreno, and A. Valls. Intelligent tourism recommender systems: A survey. *Expert Systems with Applications*, 41(16):7370–7389, 2014.

[11] J. Borràs, A. Valls, A. Moreno, and D. Isern. Ontology-based management of uncertain preferences in user profiles. In *Communications in Computer and Information Science*, pages 127–136. Springer Berlin Heidelberg, 2012.

[12] D. Bouyssou, T. Marchant, M. Pirlot, A. Tsoukias, and P. Vincke. *Evaluation and Decision Models with Multiple Criteria*, volume 86. Kluwer Academic Publishers, 2006.

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

134   **Bibliography**

[13] J. Brans. L'ingénièrie de la décision; Elaboration d'instruments d'aide à la décision. La méthode PROMETHEE. In R. Nadeau and M. Landry, editors, *L'aide à la décision: Nature, Instruments et Perspectives d'Avenir*, pages 183–213, Québec, Canada, 1982. Presses de l'Université Laval.

[14] H. Cai, M. Wang, A. Elgowainy, and J. Han. Updated greenhouse gas and criteria air pollutant emission factors and their probability distribution functions for electricity generating units. Technical report, Report of Argonne National Laboratory, U.S. Department of Energy, Chicago, Illinois, 2012.

[15] T. Calvo, G. Mayor, and R. Mesiar, editors. *Aggregation Operators*. Physica-Verlag Heidelberg, 2002.

[16] I. Cantador, A. Bellogín, and P. Castells. A multilayer ontology-based hybrid recommendation model. *AI Communications: Special Issue on Recommender Systems*, 21(2-3):203–210, 2008.

[17] C. Carlsson, M. Brunelli, and J. Mezei. Decision making with a fuzzy ontology. *Soft Computing*, 16(7):1143–1152, 2011.

[18] D. Carvalho, A. Rocha, M. Gómez-Gesteira, and C. Santos. A sensitivity study of the WRF model in wind simulation for an area of high wind energy. *Environmental Modelling & Software*, 33:23–34, 2012.

[19] F. Cena, S. Likavec, and F. Osborne. Anisotropic propagation of user interests in ontology-based user models. *Information Sciences*, 250(20):40–60, 2013.

[20] S. Chakhar and V. Mousseau. Multicriteria decision making, spatial. In *Encyclopedia of GIS*, pages 747–753. Springer US, 2008.

[21] A. I. Chatzimouratidis and P. A. Pilavachi. Sensitivity analysis of the evaluation of power plants impact on the living standard using the analytic hierarchy process. *Energy Conversion and Management*, 49(12):3599–3611, 2008.

[22] M. C. Claudia Leacock. Combining local context and wordnet similarity for word sense identification. *WordNet: An electronic lexical database*, 49(2):265–283, 1998.

[23] V. Codina and L. Ceccaroni. Taking advantage of semantics in recommendation systems. In *Proceedings of the 2010 Conference on Artificial Intelligence Research and Development: Proceedings of the 13th International Conference of the Catalan Association for Artificial Intelligence*, pages 163–172, Amsterdam, The Netherlands, The Netherlands, 2010. IOS Press.

[24] M. Darabi and N. Tabrizi. An ontology-based framework to model user interests. In *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, USA, 2016. IEEE.

[25] L. Del Vasto-Terrientes, A. Valls, R. Slowinski, and P. Zielniewicz. ELECTRE-III-h: An outranking-based decision aiding method for hierarchically structured criteria. *Expert Systems with Applications*, 42(11):4910–4926, 2015.

[26] L. Del Vasto-Terrientes, A. Valls, P. Zielniewicz, and J. Borràs. A hierarchical multi-criteria sorting approach for recommender systems. *Journal of Intelligent Information Systems*, 46(2):313–346, 2016.

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

Bibliography 135

[27] O. Edenhofer, R. Pichs-Madruga, Y. Sokona, K. Seyboth, P. Eickemeier, P. Matschoss, and C. Von Stechow. Ipcc, 2011: summary for policymakers. *IPCC Special Report on Renewable Energy Sources and Climate Change Mitigation*, 2011.

[28] J. R. Figueira, S. Greco, and M. Ehrogott. *Multiple Criteria Decision Analysis: State of the Art Surveys.* Springer New York, 2005.

[29] J. R. Figueira, S. Greco, B. Roy, and R. Słowiński. An overview of electre methods and their recent extensions. *Journal of Multi-Criteria Decision Analysis*, 20(1-2):61–85, 2013.

[30] A. García-Crespo, J. Chamizo, I. Rivera, M. Mencke, R. Colomo-Palacios, and J. M. Gómez-Berbís. SPETA: Social pervasive e-tourism advisor. *Telematics and Informatics*, 26(3):306–315, 2009.

[31] H. A. Garni, A. Kassem, A. Awasthi, D. Komljenovic, and K. Al-Haddad. A multicriteria decision making approach for evaluating renewable power generation sources in saudi arabia. *Sustainable Energy Technologies and Assessments*, 16:137–150, 2016.

[32] E. Georgopoulou, D. Lalas, and L. Papagiannakis. A multicriteria decision aid approach for energy planning problems: The case of renewable energy option. *European Journal of Operational Research*, 103(1):38–54, 1997.

[33] K. Gibert, A. Valls, and M. Batet. Introducing semantic variables in mixed distance measures: Impact on hierarchical clustering. *Knowledge and Information Systems*, 40(3):559–593, 2013.

[34] K. Govindan and M. B. Jepsen. ELECTRE: A comprehensive literature review on methodologies and applications. *European Journal of Operational Research*, 250(1):1–29, 2016.

[35] D. A. Haralambopoulos and H. Polatidis. Renewable energy projects: structuring a multi-criteria group decision-making framework. *Renewable Energy*, 28(6):961–973, 2003.

[36] A. Hawalah and M. Fasli. Utilizing contextual ontological user profiles for personalized recommendations. *Expert Systems with Applications*, 41(10):4777–4797, 2014.

[37] F. Hdioud, B. Frikh, and B. Ouhbi. Multi-criteria recommender systems based on multi-attribute decision making. In *Proceedings of International Conference on Information Integration and Web-based Applications & Services - IIWAS13*. ACM Press, 2013.

[38] S. Hong, C. J. Bradshaw, and B. W. Brook. Evaluating options for the future energy mix of japan after the fukushima nuclear crisis. *Energy Policy 56(March 2011)*, 56:418–424, 2013.

[39] J. D. Hunt, R. Bañares-Alcántara, and D. Hanbury. A new integrated tool for complex decision making: Application to the UK energy sector. *Decision Support Systems*, 54(3):1427–1441, 2013.

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

136   **Bibliography**

[40] A. Ishizaka and P. Nemery. *Multi-Criteria Decision Analysis: methods and software.* John Wiley & Sons, 2013.

[41] G. Jakus, V. Milutinović, S. Omerović, and S. Tomažič. *Concepts, Ontologies, and Knowledge Representation.* Springer-Verlag, 2013.

[42] X. Jiang and A.-H. Tan. Learning and inferencing in user ontology for personalized semantic web search. *Information Sciences*, 179(16):2794–2808, 2009.

[43] J. K. Kaldellis, A. Anestis, and I. Koronaki. Strategic planning in the electricity generation sector through the development of an integrated delphi-based multi-criteria evaluation model. *Fuel*, 106:212–218, 2012.

[44] R. L. Keeney and H. Raiffa. *Decisions with multiple objectives: preferences and value tradeoffs.* J. Wiley, New York, 1976.

[45] R. L. Keeney and H. Raiffa. *Decisions with Multiple Objectives.* Cambridge University Press, 1993.

[46] S. Köppen, H. Fehrenbach, S. Markwardt, and A. Hennecke. Implementing the GBEP indicators for sustainable bioenergy in germany. Technical report, Report from IINAS and IFEU, Heidelberg, Darmstadt, Berlin, Germany., 2014.

[47] D. Küçük and Y. Arslan. Semi-automatic construction of a domain ontology for wind energy using wikipedia articles. *Renewable Energy*, 62:484–489, 2014.

[48] W. Kunz and H. Rittel. Issues as elements of information systems. Technical report, Center for Planning and Development Research, University of California at Berkeley, 1970.

[49] C. Lamsfus, C. Grün, A. Alzua-Sorzabal, and H. Werthner. Context-based match-making to enhance tourists' experience. *Journal for the Informatics Professional*, 203:17–23, 2010.

[50] Y. Li, Z. Bandar, and D. McLean. An approach for measuring semantic similarity between words using multiple information sources. *IEEE Transactions on Knowledge and Data Engineering*, 15(4):871–882, 2003.

[51] A. Luberg, T. Tammet, and P. Jrv. Smart city: A rule-based tourist recommendation system. In *Information and Communication Technologies in Tourism 2011*, pages 51–62. Springer, Vienna, 2011.

[52] J. Macknick, R. Newmark, G. Heath, and K. C. Hallett. Review of operational water consumption and withdrawal factors for electricity generating technologies. Technical report, National Renewable Energy Laboratory, U.S. Department of Energy, Golden, Colorado, 2011.

[53] M. Martínez-García, A. Valls, and A. Moreno. Construction of an outranking relation based on semantic criteria with ELECTRE-III. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems Part II*, pages 238–249. Springer International Publishing, 2016.

[54] M. Martínez-García, A. Valls, A. Moreno, and A. Aldea. A semantic multi-criteria approach to evaluate different types of energy generation technologies. *Environmental Modelling and Software, Special Issue: Environmental Data Science*, 2018.

[55] J. Meldrum, S. Nettles-Anderson, G. Heath, and J. Macknick. Life cycle water use for electricity generation: a review and harmonization of literature estimates. *Environmental Research Letters*, 8(1):015031, 2013.

[56] E. G. Michael Doumpos. *Multicriteria Decision Aid and Artificial Intelligence: Links, Theory and Applications.* 2013.

[57] S. E. Middleton, D. De Roure, and N. R. Shadbolt. *Ontology-based Recommender Systems*, pages 477–498. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[58] S. E. Middleton, N. R. Shadbolt, and D. C. D. Roure. Ontological user profiling in recommender systems. *ACM Transactions on Information Systems*, 22(1):54–88, 2004.

[59] P. L. Mínguez I, Berrueta D. CRUZAR: An application of semantic matchmaking to e-tourism. *Cases on Semantic Interoperability for Information Systems Integration: Practices and Applications*, pages 255–271, 2010.

[60] A. Moreno, A. Valls, D. Isern, L. Marin, and J. Borràs. SigTur/e-destination: Ontology-based personalized recommendation of tourism and leisure activities. *Engineering Applications of Artificial Intelligence*, 26(1):633–651, 2013.

[61] A. S. Niaraki and K. Kim. Ontology based personalized route planning system using a multi-criteria decision making approach. *Expert Systems with Applications*, 36(2):2250–2259, 2009.

[62] T. D. Noia, V. C. Ostuni, P. Tomeo, and E. D. Sciascio. SPrank: Semantic path-based ranking for Top-N recommendations using linked open data. *ACM Transactions on Intelligent Systems and Technology*, 8(1):1–34, 2016.

[63] A. Papadopoulos and A. Karagiannidis. Application of the multi-criteria analysis method Electre III for the optimisation of decentralised energy systems. *Omega*, 36(5):766–776, 2008.

[64] H. Parsaei, M. Wilhelm, and S. Kolli. Application of outranking methods to economic and financial justification of CIM systems. *Computers and Industrial Engineering*, 25(1-4):357–360, 1993.

[65] E. G. M. Petrakis, G. Varelas, A. Hliaoutakis, and P. Raftopoulou. X-similarity: Computing semantic similarity between concepts from different ontologies. *Journal of Digital Information Management*, 4:233–237, 2006.

[66] G. Pirró and N. Seco. Design, implementation and evaluation of a new semantic similarity metric combining features and intrinsic information content. In *On the Move to Meaningful Internet Systems: OTM 2008*, pages 1271–1288. Springer, Monterrey, México, 2008.

[67] G. R and D. H. Lapproche multidcideur multicritre daide la dcision. Technical report, Institute de Technologie dIndonsie, 2000.

[68] R. Rada, H. Mili, E. Bicknell, and M. Blettner. Development and application of a metric on semantic nets. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(1):17–30, 1989.

[69] P. Resnik. Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, IJCAI'95, pages 448–453, Montreal, Quebec, Canada, 1995. Morgan Kaufmann Publishers Inc.

[70] F. Ribeiro, P. Ferreira, and M. Araújo. Evaluating future scenarios for the power generation sector using a multi-criteria decision analysis (MCDA) tool: The portuguese case. *Energy*, 52:126–136, 2013.

[71] M. Rogers and M. Bruen. Choosing realistic values of indifference, preference and veto thresholds for use with environmental criteria within ELECTRE. *European Journal of Operational Research*, 107(3):542–551, 1998.

[72] E. L. L. Rovere, J. B. Soares, L. B. Oliveira, and T. Lauria. Sustainable expansion of electricity sector: Sustainability indicators as an instrument to support decision making. *Renewable and Sustainable Energy Reviews*, 14(1):422–429, 2010.

[73] B. Roy. Critères multiples et modélisation des préférences (l'apport des relations de surclassement). *Revue d'économie politique*, 84(1):1–44, 1974.

[74] B. Roy and D. Bouyssou. *Aide Multicritère la Décision : Méthodes et Cas*. Economica, 1993.

[75] M. Ruiz-Montiel and J. F. Aldana-Montes. Semantically enhanced recommender systems. In *Lecture Notes in Computer Science*, pages 604–609. Springer Berlin Heidelberg, 2009.

[76] V. Schickel-Zuber and B. Faltings. Overcoming incomplete user models in recommendation systems via an ontology. In *Advances in Web Mining and Web Usage Analysis*, pages 39–57. Springer Berlin Heidelberg, 2006.

[77] S. Sendhilkumar and T. V. Geetha. Personalized ontology for web search personalization. In *Proceedings of the 1st Bangalore annual Compute conference*, pages 18:1–18:7, New York, NY, 2008. ACM.

[78] S. E. Shmelev and J. C. V. D. Bergh. Optimal diversity of renewable energy alternatives under multiple criteria: An application to the UK. *Renewable and Sustainable Energy Reviews*, 60:679–691, 2016.

[79] P. Shoval, V. Maidel, and B. Shapira. An Ontology- Content-based Filtering Method. *International Journal of Information Theories and Applications*, 15(4):51–63, 2008.

[80] S. J. B. Shum, A. M. Selvin, M. Sierhuis, J. C. an Charles B. Haley, and B. Nuseibeh. Hypermedia support for argumentation-based rationale: 15 years on from gibis and qoc. *Rationale Management in Software Engineering*, pages 111–132, 2006.

[81] A. Sieg, B. Mobasher, and R. Burke. Web search personalization with ontological user profiles. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 525–534, New York, NY, 2007. ACM.

[82] S. Stagl. Multicriteria evaluation and public participation: the case of UK energy policy. *Land Use Policy*, 23(1):53–62, 2006.

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

Bibliography 139

[83] L. Stamford and A. Azapagic. Life cycle sustainability assessment of UK electricity scenarios to 2070. *Energy for Sustainable Development*, 23:194–211, 2014.

[84] E. W. Stein. A comprehensive multi-criteria model to rank electric energy production technologies. *Renewable and Sustainable Energy Reviews*, 22:640–654, 2013.

[85] E. Strantzali and K. Aravossis. Decision making in renewable energy investments: A review. *Renewable and Sustainable Energy Reviews*, 55:885–898, 2016.

[86] Z. Su, J. Yan, H. Ling, and H. Chen. Research on personalized recommendation algorithm based on ontological user interest model. *Journal of Computational Information Systems*, 8(1):169–181, 2012.

[87] K. ten Hagen, R. Kramer, M. Hermkes, B. Schumann, and P. Mueller. Semantic matching and heuristic search for a dynamic tour guide. In *Information and Communication Technologies in Tourism 2005*, pages 149–159. Springer, Vienna, 2005.

[88] V. Torra. The weighted owa operator. *International Journal of Intelligent Systems*, 12(2):153–166, 1997.

[89] V. Torra and Y. Narukawa. *Modeling Decisions*. Springer Berlin Heidelberg, 2007.

[90] V. Torra and Y. Narukawa. A view of averaging aggregation operators. *IEEE Transactions on Fuzzy Systems*, 15(6):1063–1067, 2007.

[91] W. U, C. I, and W. T. From social networks to distributional properties: A comparative study on computing semantic relatedness. In *Of Thirty-First Annual meeting of the Cognitive Science Society*, pages 3016–3021, Amsterdam, Netherlands, 2009.

[92] A. Valls, A. Moreno, and J. Borràs. Preference representation with ontologies. In M. Doumpos and E. Grigoroudis, editors, *Multicriteria Decision Aid and Artificial Intelligence*, pages 77–99. John Wiley & Sons, Chichester, UK, 2013.

[93] J.-J. Wang, Y.-Y. Jing, C.-F. Zhang, and J.-H. Zhao. Review on multi-criteria decision analysis aid in sustainable energy decision-making. *Renewable and Sustainable Energy Reviews*, 13(9):2263–2278, 2009.

[94] W. Wang, G. Zeng, D. Zhang, Y. Huang, Y. Qiu, and X. Wang. An intelligent ontology and bayesian network based semantic mashup for tourism. In *Proceedings IEEE Congress on Services*, pages 128–135. IEEE, 2008.

[95] Y.-M. Wang and Y. Luo. Integration of correlations with standard deviations for determining attribute weights in multiple attribute decision making. *Mathematical and Computer Modelling*, 51(1-2):1–12, 2010.

[96] Z. Wu and M. Palmer. Verbs semantics and lexical selection. In *Proceedings of the 32nd annual meeting on Association for Computational Linguistics -*, pages 133–138, Las Cruces, New Mexico, 1994. Association for Computational Linguistics.

[97] R. R. Yager. Quantifier guided aggregation using OWA operators. *International Journal of Intelligent Systems*, 11(1):49–73, 1998.

[98] R. R. Yager, J. Kacprzyk, and G. Beliakov, editors. *Recent Developments in the Ordered Weighted Averaging Operators: Theory and Practice.* Springer Berlin Heidelberg, 2011.

[99] C.-N. Ziegler, G. Lausen, and L. Schmidt-Thieme. Taxonomy-driven computation of product recommendations. In *Proceedings of the Thirteenth ACM conference on Information and knowledge management*, pages 406–415, New York, 2004. ACM Press.

UNIVERSITAT ROVIRA I VIRGILI
PROPIETARIAS A LA ESPERA: MIGRACIÓN INTERNACIONAL, HERENCIA Y GÉNERO EN DOS COMUNIDADES
INDÍGENAS OAXAQUEÑAS
María Martínez Iglesias

UNIVERSITAT
ROVIRA i VIRGILI