Chapter 1

Introduction

Large amounts of digital information are produced on a daily basis in the course of normal human activity, such as news reports, scientific presentations, meetings, etc. Documents generated from such events can be of different types in terms of the media (e.g., written, audio, video), the domain (e.g., politics, science, business), or the context of their creation (e.g., newspaper prose, telephone conversation, political speech, business meeting discussion, or scientific conference presentation). In this context, to deal with the increasing amount of available information it is helpful to use Automatic Summarization (AS). Some examples of summaries in daily activities are the production of news headlines, student note outlines, meeting minutes, film or book synopsis, and weather forecasts or stock market bulletins.

This chapter is divided in three sections, in the next one the problem to be solved is defined, Section 1.2 presents the aim of the thesis and Section 1.3 describes the research method employed.

1.1   Problem to be solved

To define the problem to be solved in this thesis, we should first ask ourselves the following questions:

What is a summary?   "no one seems to know exactly what a summary is"(Hovy 2005). For example, some generic definitions of what a summary is are presented bellow.

- (Cambridge Dictionaries Online 2007) *"a short clear description that gives the main facts or ideas about something".*

- (Encarta Dictionary Online 2007) *"short version containing gist of something: a shortened version of something that has been said or written, containing only the main points".*

- (Sparck-Jones 2001) *"a reductive transformation of a source text into a summary text by extraction or generation".*

- (Hovy 2005) *"a text that is produced from one or more texts, that contains a significant portion of the information in the original text(s), and that is no longer than half of the original text(s)". "Text" includes multimedia documents, on-line documents, hypertexts, etc. Of the many types of summary that have been identified (Sparck-Jones 1999); (Hovy and Lin 1999).*

What is Automatic Summarization? This thesis addresses the process of producing summaries by means of automatic techniques defined as follows:

- (Mani and Maybury 1999) *"to take an information source, extract content from it, and present the most important content to the user in a condensed form and in a manner sensitive to the user's or application's needs".*

The aim of the work presented in the following chapters is to propose a modular architecture able to produce summaries from documents with different characteristics. Moreover, the architecture should be parametrizable to take into account general user needs (e.g. size of the summary, output media, general content, novelty or related to a query). In this thesis the AS problem has been tackled more in breadth than in depth. Section 1.1.1 presents several facets to be taken into account when creating a multitask automatic summarizer. Some of the aspects relevant for the thesis are exemplified by citing automatic approaches. In addition, due to the fact that the evaluation process is always an important issue when developing information systems, Section 1.1.2 briefly introduces some specific issues of the AS evaluation process.

1.1.1 Automatic Summarization Aspects

Effective summarizing requires an explicit and detailed analysis of context factors, as is apparent when we recognize that what summaries should be like is defined by what they are wanted for. The parameters to be taken into account in AS systems have been widely discussed in the literature (Mani and Maybury 1999; Hovy 2001; Mani 2001). Acording to (Sparck-Jones 1999) work, three main aspects affect the process of Summarization: input, purpose and output. These aspects are described bellow.

Input Aspects

The first main aspects relevant to the AS process are the input aspects. The features of the document to be summarized crucially determine the way a summary can be obtained.

Document Structure.   Besides textual content, heterogeneous documental information concerning format can be found in a source document, for example, labels that mark headers, chapters, sections, lists, tables, etc. This information has been used by (Kan 2003) and (Elhadad et al. 2005) to exploit the organization of medical articles in sections in order to build a tree-like representation. On the other hand, (Teufel and Moens 2002) systematize the structural properties of scientific articles to assess the contribution of each textual segment.

However, it can also be the case that the information provided by the structure of the document does not help in the analysis. In this case, document structure is not taken into account.

Domain.   *Domain-sensitive* systems are only able to obtain summaries of texts that belong to a pre-determined domain, with varying degrees of portability. In contrast, *general purpose* systems are not dependant on information about domains, which usually results in a more shallow approach to the analysis of the input documents.

Nevertheless, some general purpose systems are prepared to exploit domain specific information. For example, the meta summarizer developed at Columbia University (Barzilay et al. 1999; Barzilay et al. 2001; Hatzivassiloglou et al. 1999; Hatzivassiloglou et al. 2001; McKeown et al. 2002) applies different summarizers for different kinds of documents: MULTIGEN (Barzilay et al. 1999; McKeown et al. 1999) is specialized in simple events; DEMS (Schiffman et al. 2001) (with the bio configuration) deals with biographies; and for the rest of documents, DEMS has a default configuration that can be resorted to.

In general, working with domain-specific documents alleviates some NLP problems, such as restricting the valid senses for a word, but new problems appear, such as terminological units, jargon, etc.

Specialization level.   A text may be broadly characterized as *ordinary*, *specialized*, or *restricted*, in relation to the presumed subject knowledge of the source text readers. This aspect can be considered the same as the *domain* aspect discussed above.

Restriction on the language.   The language of the input can be *general* language or *restricted to a sublanguage* within a domain, purpose or audience. It may be sometimes necessary to preserve the sublanguage in the summary.

Scale.   Different summarizing strategies have to be adopted to handle different text lengths. Indeed, the analysis of the input text can be performed at different granularities, for example, when determining meaning units. In the case of news articles, sentences or even clauses are usually considered the minimal units, whereas for longer documents, like reports or books, paragraphs seem a more adequate . The techniques for segmenting the input text in meaning units differ depending on the length of the document: for shorter texts, orthography and syntax, even discourse boundaries (Marcu 1997a) indicate significant boundaries, for longer texts, topic segmentation (Kozima 1993; Hearst 1994) is more usual. There is a growing interest on defining such minimal meaning units for both summarizing and evaluating summaries. The Summary Content Unit (SCU), defined in the framework of the pyramid method (Nenkova and Passonneau

2004) and Basic Element (BE) (Hovy et al. 2005) go in the direction of defining such minimal TU.

Media. Although the main focus of summarization is *textual* summarization, in recent years summaries of *non-textual documents*, like videos, meeting records, images, graphics or tables have also been undertaken. Sometimes this non-textual documents are transformed to text by means of an Automatic Speech Recognizer (ASR) or an Optical Character Recognizer (OCR). The result of this transformation is usually an ill-formed text.

Two broad cases can be distinguished, cases where the source input and the summary output is in the same media, and cases where the source is in one media and the summary in the other. An example of cross-media work is (Takeshita et al. 1997), that select representative images from video by analyzing the topic structure of the closed-captioned text. (Futrelle 1999) summarizes diagrams from scientific and technical text. (Merlino and Maybury 1999) propose a tool for searching, browsing and summarizing TV news broadcasts by using information from the audio, the video and the text using a variety of presentation techniques.

The complexity of multimedia summarization has prevented the development of wide coverage systems, which means that most summarization systems that can handle multimedia information are limited to specific domains or textual genres . However, some research efforts try to integrate information from different media by providing a suite of tools for describing semantics relations of multimedia (Benitez and Chang 2002).

Genre. Some systems exploit typical genre-determined characteristics of texts, such as the pyramidal organization of newspaper articles, or the argumentative development of a scientific article. Some example of *genre-dependent* systems are: (Elhadad et al. 2005), in healthcare; (Kan 2003), in medical articles; (McKeown and Radev 1995), in agency news; (Hauptmann and Witbrock 1997), in broadcast fragments; (Zechner 2001), in meeting recording; (Muresan et al. 2001), in e-mails; and (Radev et al. 2002), in web pages. In contrast, those systems that do not exploit the genre characteristics of texts are *genre-independent*.

Unit. The input to the summarization process can be a *single document* or *multiple documents*, either simple text or multimedia information such as audio or video (Sundaram 2002).

Language. Systems can be *language-independent*, *crosslingual*, when exploiting characteristics of documents that hold cross-linguistically (Radev et al. 2003; Pardo et al. 2003), or otherwise their architecture can be determined by the features of a concrete language, being then *language-dependent*. This means that some adaptations must be carried out in the system to deal with different languages. This aspect is extended in Section 2.2.2.

Purpose Aspects

The second main aspect set of the AS process are the purpose aspects, which deal with the specific characteristics of the task or the final utility of the summary.

Situation.   Summarization systems can perform *general* summarization or they can be embedded in larger system, as an intermediate step for another NLP task or application, like Machine Translation (MT),Information Retrieval (IR) or Question & Answering (QA). *Task-driven* summarization presents the advantage that systems can be evaluated with respect to the improvement they introduce in the final task they are applied to.

Audience.   In case a user profile is available, summaries can be adapted to the needs of specific users, for example, the user's prior knowledge on a determined subject. *Background* summaries assume that the reader's prior knowledge is poor, and so extensive information is supplied, while *just-the-news* are those kind of summaries conveying only the newest information on an already known subject. *Briefings* are a particular case of the latter, they collect representative information from a set of related documents.

Usage.   Summaries can be sensitive to determined uses: retrieving source text (Kan et al. 2001), previewing a text (Leuski et al. 2003), refreshing the memory of an already read text, sorting, etc.

Output Aspects

The third important AS relevant aspect set are those related with the features of the document to be presented as output.

Content.   A summary may try to represent all relevant features of a source text or it may focus on some specific ones, which can be determined by queries, subjects, etc. *Generic* summaries are text-driven, while *user-focused* (or query-driven) ones rely on a specification of the user's information need, often expressed by a question or a list of keywords.

Format.   The output of a summarization system can be plain text or it can be formatted. Formatting can be targeted to many purposes: conforming to a pre-determined style (tags, organization in fields), improving readability (division in sections, highlighting), or classifying the relationship between themes and documents (Ando et al. 2000).

Style.   A summary can be *informative*, if it covers the topics in the source text; *indicative*, if it provides a brief survey of the topics addressed in the original; *aggregative*, if it supplies information not present in the source text that completes some of its information or elicits some hidden information (Teufel and Moens 2002); or *critical*, if it provides an additional valoration of the summarized text.

Production Process.   The resulting summary text can be an *extract*, if it is composed by literal fragments of text, or an *abstract*, if it is generated. The type of summary output desired can be relatively polished, for example, if text is well-formed and connected, or else more fragmentary in nature (e.g., a list of key words). A very aggressive form of summarization,

known as very short summaries, ultrasummarization or headline generation, usually results in unconnected lists of terms or short sequences of terms.

There are intermediate options, mostly concerning the nature of the fragments (TUs) that compose extracts, which can range from topic-like passages, paragraph or multiparagraph long, to clauses or even phrases. In addition, some approaches perform editing operations in the summary, overcoming the incoherence and redundancy often found in extracts, but at the same time avoiding the high cost of a NLG system. (Jing and McKeown 2000) apply six re-writing strategies to improve the general quality of an extract-based summary by edition operations like deletion, completion or substitution of clausal constituents.

Surrogation. Summaries can stand in place of the source as a surrogate, or they can be linked to the source (Kan et al. 2001; Leuski et al. 2003), or even be presented in the context of the source, e.g., by highlighting source text, (Lehmam and Bouvet 2001).

Length. The targeted length of the summary crucially affects the informativeness of the final result. This length can be determined by a compression rate, that is to say, a ratio of the summary length with respect to the length of the original text. Summary length can also be determined by the physical context where the summary is to be displayed. For example, in the case of delivery of news of summaries to hand-helds (Boguraev et al. 2001; Buyukkokten et al. 2001; Corston-Oliver 2001), the size of the screen imposes severe restrictions to the length of the summary. Headline generation is another application where the length of summaries is clearly determined (Witbrock and Mittal 1999; Daumé III et al. 2002). In very short summaries, coherence is usually sacrificed to informativeness, so lists of words are considered acceptable (Kraaij et al. 2002; Zajic et al. 2002).

1.1.2   Automatic Summarization Evaluation

The evaluation of a Natural Language Processing (NLP) system (or some of its components) is complex, controversial and costly. Recently, it has also become a critical issue for many difficult NLP applications, such as AS.

According to (Sparck-Jones and Galliers 1999), two major types of NLP evaluation can be differentiated: *intrinsic* and *extrinsic* evaluations. Both can be applied to evaluate AS systems. The first type directly measures the quality of the created summary by evaluating some intrinsic properties, such as coverage or readability. The second type of evaluation indirectly measures summary performance in a task dependant on the quality of the summary. Usually, extrinsic methods are used to compare the accuracy of different systems without assigning absolute scores to them. TIPSTER SUMMAC (Mani et al. 1998) and the Japanese NII Test Collection for IR Systems (NTCIR)[1] (NTCIR 2006) are good examples of AS extrinsic evaluations. In contrast,

---

[1]`http://research.nii.ac.jp/ntcir/index-en.html`

the ones carried out by the National Institute of Standards and Technology (NIST) in the DUC[2] contest are good examples of AS intrinsic evaluations.

Evaluating summaries, either manually or automatically, is a hard task. The main difficulty in evaluation comes from the impossibility of building a fair gold standard against which the results of the system we wish to evaluate can be compared. This difficulty is due to the very low agreement among human evaluators when faced with interpretative decisions. This lack of agreement comes, in turn, from the difficulty of defining the set of measurable properties that contribute to the quality of summaries. Usually, different properties are taken into account and it is a difficult task to select the most appropriate ones and to combine them properly. (Mani 2001) provides a clear picture of summary evaluation, both by human judges and by automated metrics.

Different initiatives have been taken to try to solve the summary evaluation problem trying to provide the most neutral and objective evaluation mechanisms. These mechanisms include:

- The organization of evaluation contests to provide training and test material, objective metrics and evaluation criteria, human judgment of the performance of the systems, etc. As said before, among these contests we can find DUC, that since 2001 has been the major international forum for comparing summarization systems. Participating groups work on the same tasks, use the same collection of documents and metrics, and then compare their results.

- The production of annotated corpora to be used as gold standards such as the human built extracts obtained from the DUC 2001 data (Conroy et al. 2001) or the CAST project corpus (Orasan et al. 2003), containing news articles taken from the Reuters Corpus and a few science texts from the British National Corpus, as well as summary sentences marked as essential or important.

- The study of the inter-assessor agreement with measures such as Kappa (Carletta 1996). Although Kappa has been widely used as a measure of agreement between evaluators, recent studies point out some limitations and propose alternatives. See Chapter 4 of (Boleda 2007) for an excellent survey of human judgment agreement measures.

- The use of automatic evaluation metrics, including well known simple metrics, like precision and recall, or the ones based on n-gram overlap as Recall-Oriented Understudy for Gisting Evaluation (ROUGE) (Lin 2004).

- The combination of different metrics and judgments to increase the robustness of the evaluation process using evaluation frameworks such as QARLA(Amigó et al. 2005; Amigó 2006).

---

[2]http:/www-nlpir.nist.gov/projects/duc/

An important issue in the evaluation process is to establish carefully a real summarization task. Tables 1.1, 1.2 and 1.3 show summary examples for different summarization tasks tackled along this thesis. For each task, a table with manual and automatic summaries (or "peers") are presented. Manual summaries are usually considered as a reference to evaluate the content of the automatic ones.

Manual Summary	
Paper KeyWords	excitation vocoder text-to-speech synthesis
Paper Title	a new model of excitation for text-to-speech synthesis
Human 1	distance text to speech synthesis excitation model residual signal
Human 2	new excitation model for the distance speech synthesis residual signal
Automatic Summary	
	excitation model for the distance speech synthesis quality of text

Table 1.1: Example of 10-word summaries produced from an automatic oral presentation transcript. The human summaries and the corresponding title and the keywords of the paper are used as reference and summaries.

In the first example, Table 1.1, the objective of the task is to produce very short summaries from an oral presentation transcript, in other words, to briefly present the main subject of the presentation. The resulting summary will be displayed in a hand-held, or the corresponding audio segment will be reproduced, this task is treated in Chapter 5. In Table 1.1, aside from the manual summaries produced by the assessors, the title and the keywords of the scientific paper related to the oral presentation to be summarized are also considered as manual summaries to be used in the evaluation of the automatic summary content. In the other examples, Tables 1.2 and 1.3, the goal is to summarize a set of related documents trying to give answer to a user need (or "query"). These examples try to illustrate that different aspects can be considered for the evaluation of each task. In Table 1.2 the user need is represented as the topic of the set of documents (oral transcript, scientific papers, slides, author notes, ...) from the oral presentation to be summarized and a list of keywords to be taken into account as a query to produce the summary (see Chapter 7 for more details). In Table 1.3 the user need is represented by a title describing the cluster of related documents to be summarized, a narrative description of the information that should be answered using the cluster of related documents, and the summary content granularity (general or specific). The expected summary granularity indicates user/task context information for systems and human summarizers. While general summaries are supposed to capture some general user/task preference in a simple user profile, specific summaries explicitly reflect the specific interests of a potential user in a task context. This task is one of the recently proposed by DUC contest, more information about it can be found in

User need
topic: Speech Translation
query: Statistical Machine Translation, Noisy Channel Paradigm
Manual Summaries
The translation model used is the CMU Statistical Machine Translation toolkit. The one best sentence inputted is converted into a single path lattice first. Statistical machine translation work is based on the noisy channel paradigm. The noisy-channel approach is a basic concept of SMT that was adapted to the problem of disfluency cleaning, by assuming that "clean" i.e. fluent speech gets passed through a noisy channel. The channel adds noise to the clean speech and creates "noisy" disfluent speech as output. The modelling task is to build statistical models which capture the characteristics of the translation and the target language.
In conventional speech translation systems, the recognizer outputs a single hypothesis which is then translated by the Statistical Machine translation (SMT) system. This approach has the limitation of being largely dependent on the word error rate. A translated sentence from English to French is garbled or is encoded by some noisy channel and we need to model this noisy channel in some statistically meaningful way. The cleaning component is based on a noisy channel approach, a basic concept of SMT that is adapted to the problem of disfluency cleaning, by assuming that fluent speech gets passed through a noisy channel.
While speech recognition emerged to be rapidly adaptable to new languages in large domains, translation still suffer from lack of large parallel corpora for statistical machine translation. Statistical machine translation work is based on the noisy channel paradigm. The model to describe this noisy channel needs to provide us with a probability of an English target sentence given a French source sentence. In SMT we use alignments to establish correspondences between the positions of the source and the target sentences. In conventional speech translation systems, the recognizer outputs a single hypothesis which is then translated by the SMT system.
Automatic Summary
Furthermore, while speech recognition emerged to be rapidly adaptable to new languages in large domains, translation still suffer from the need of hand-crafted grammars for interlinguabased approaches or the lack of large parallel corpora for statistical machine translation. The translation model used is the CMU Statistical Machine Translation toolkit CMU-SMT 13. um it is based on the noisy channel paradigm.The channel adds noise to the clean speech and creates " noisy ", i.e. disfluent speech as output. Given a " noisy " string N the goal is to recover the " clean " string such that p N becomes

Table 1.2: Given a user query, example of 100-word manual and automatic summaries produced from different sorts of documents from an oral presentation.

User need
topic: d426a
title: Welsh devolution and British Parliament
narrative: What is the status of British plans for re-structuring the Welsh government? How is the issue of separation from British rule being treated by Welsh and British legistators?
granularity: general

Manual Summaries
Wales is following Scotland, and moving towards a call for an elected assembly with devolved powers, as advocated by the Labour Party. Labour has committed to the creation of a Welsh assembly, and party leader Tony Blair set out proposals for devolution, setting off a constitutional battle with the Tories. Conservatives oppose any form of devolution, and want to maintain a strong Welsh Office with a cabinet minister, believing that would produce the best results for Wales. Prime Minister John Major and the Tories are against the establishment of a Welsh parliament, which has eroded the usual support conservative legislators had received in Wales. Plaid Cymru, the Welsh nationalist party, stepped up its campaign for equal rights to Welsh self-determination, demanding equal constitutional treatment with Northern Ireland, and Scotland. The British government is pressing ahead with plans to reform the structure of local government in Wales. It will establish an elected Welsh assembly, with law-making and financial powers, to replace the current two-tier system of county and district councils with single-purpose, unitary authorities. The government intends to set up 21 new authorities to replace the eight counties and 37 districts in Wales. Shadow elections to the new unitary authorities will be held as early as next year. Implementation of the local government reform will take place in April 1995.
On December 31, 1994 a fight was expected to begin on the devolution of parliamentary powers to Wales.Tony Blair, the Labor party leader, planned to pursue his intentions to set up a Welsh assembly. The Conservative party leader, John Major, called Blair's plan the most dangerous propositions ever put before the British nation. The most recent push for self-rule for Wales became strong in 1992. White papers recommended government reforms, which replaced the eight county councils and 37 district councils with 21 unitary authorities. Plans were to create shadow councils for Wales and have them operational by the spring of 1995. In 1993 these plans were delayed for one year with shadow council elections scheduled for 1995 and in place in 1996. Support for the devolution of Wales generally followed party lines. In 1992 Conservatives called the present arrangement the best for the UK and even the Welch minister, Alun Michael, did not consider devolution an issue. Plaid Cymru, the nationalist party of Wales, called for a Welsh parliament and wanted full self-government for Wales within the European Community. By 1994 Plaid Cymru demanded equal rights with Scotland for self-determination. Labor party leaders continued to back regional administrations for Wales and an elected assembly, although the 1992 party leader Neil Kinnock said the working class would gain nothing from the changes. The Liberal-Democratic party leaders in 1992 believed there was popular support for devolution and in 1994 minister Paddy Ashdown said he possibly would support the assembly.

Automatic Summary
The manifesto, titled Towards 2000, says any funding from Europe must be handled within Wales by a Welsh parliament, and must be in addition to money from the UK government. Plaid's three seats in the last parliament had all been won in the one Welsh county where Welsh is very much the medium of conversation. PLAID CYMRU The Welsh nationalist party is highly unlikely to make any deal with the Tories and says any deal will not support any government without a promise of a Welsh parliament within a year.' their mad policies have infiltrated the Welsh party and are pushing it in a direction and at a rate that the people of Wales do not altogether want'. Conservatives are adamantly opposed to any form of devolution. Accepting the government's position that reform should be cost-neutral the CBI claims the handling of strategic issues could be fragmented if the number of authorities were to be too large and that the small Welsh authorities ' may also be at a disadvantage compared to the English counties. ' The calls for an elected assembly come mainly from Labour, the dominant political force in Wales. The government is facing growing demands from Welsh and Scottish nationalists for equal constitutional treatment with Northern Ireland, including the establishment of devolved assemblies. Nationalists in both countries see the Ulster peace talks as the trigger for a campaign to highlight the contradictions between the government's approach to Northern

Table 1.3: Given a user need, example of 250-word manual and automatic summaries produced from a set of 25-50 documents.

Chapter 6. Human and automatic summarizers have to create a 150-word summary (in Table 1.2) and a 250-word summary (in Table 1.3) from a cluster of related documents satisfying a user need. The manual summaries (3 in Table 1.2 and 2 in Table 1.3) will be used as reference to evaluate the summary content.

1.2   Aim of the Thesis

The final goal of this thesis is to provide a flexible multitask summarizer architecture to allow a simple integration of different techniques, tools, and resources. This goal can be decomposed in three main research purposes:

First, to study different AS aspects and existing systems to propose a modular architecture to facilitate the re-usability of code for different tasks. The architecture should permit:

- To process different sorts of input documents in different languages.

- To create different sorts of summaries depending on the summarization task to be performed or the user information need.

- To create different content summaries: generic text-driven or query-driven summaries.

- To produce different output formats: synthesized voice, audio/video or plain text segments.

- To allow processing raw text coming from different media, domain, or genre.

Second, to instantiate some prototypes within this architecture for different summarizing tasks. These tasks include:

- To summarize documents in Catalan, Spanish, and English.

- To consider text and/or audio transcribed documents as input.

- To process news or scientific presentation documents.

- To deal with well written text or ill-formed text as input.

- To process a variable number of documents. The number of the input units to be summarized can range from 1 to about 50 documents.

- To produce different length summaries. The length of the output can range from 10-word for headline production to 250-word.

- To give answer to different types of user needs (a complex question or a list of keywords), when the content of the summary is query-focused.

- To study the question of porting the implemented approaches to other tasks, languages, domains or media.

Third, to create an evaluation framework to analyze the performance of several approaches in the domains of written news and scientific oral presentations. In this thesis, we mainly focus in intrinsic evaluation:

- The participation in the international contests to obtain manual (human) evaluations for our prototypes.

- The creation of summary reference corpora and the definition of procedures to evaluate the summary content produced for different tasks.

1.3 Research Method

Figure 1.1 represents the circular method applied during the research process to obtain the thesis presented in this document. Starting by studying the existing systems and summarization methodologies, a prototype based design is used and some approaches are implemented applying different summarization techniques. Prototypes have been extremely useful in understanding what the software is supposed to do. To be able to evaluate and improve the implemented approaches, several *in-house* (*own* in Figure 1.1) evaluations have been proposed. Furthermore, when possible, the implemented approaches have been evaluated in *internationally* recognized evaluation frameworks. An important requirement in the generic architecture design process was to facilitate the re-use of existing software; new approaches should be able to easily re-use architecture components previously instantiated by other approaches. Moreover, this generic architecture should allow to instantiate any technique used in existing systems or methodologies.

1.4 Document Outline

The thesis is divided into eight chapters. The first chapter is this Introduction, and the remaining chapters are organized as follows:

- Chapter 2 presents an overview of the state-of-the-art in Automatic Summarization.

- Chapter 3 provides an overview of the general architecture proposed in this thesis.

- Chapter 4 discusses the experimental results obtained in Multilingual summarization.

- Chapter 5 details the experiments carried out in Spontaneous Speech Summarization.
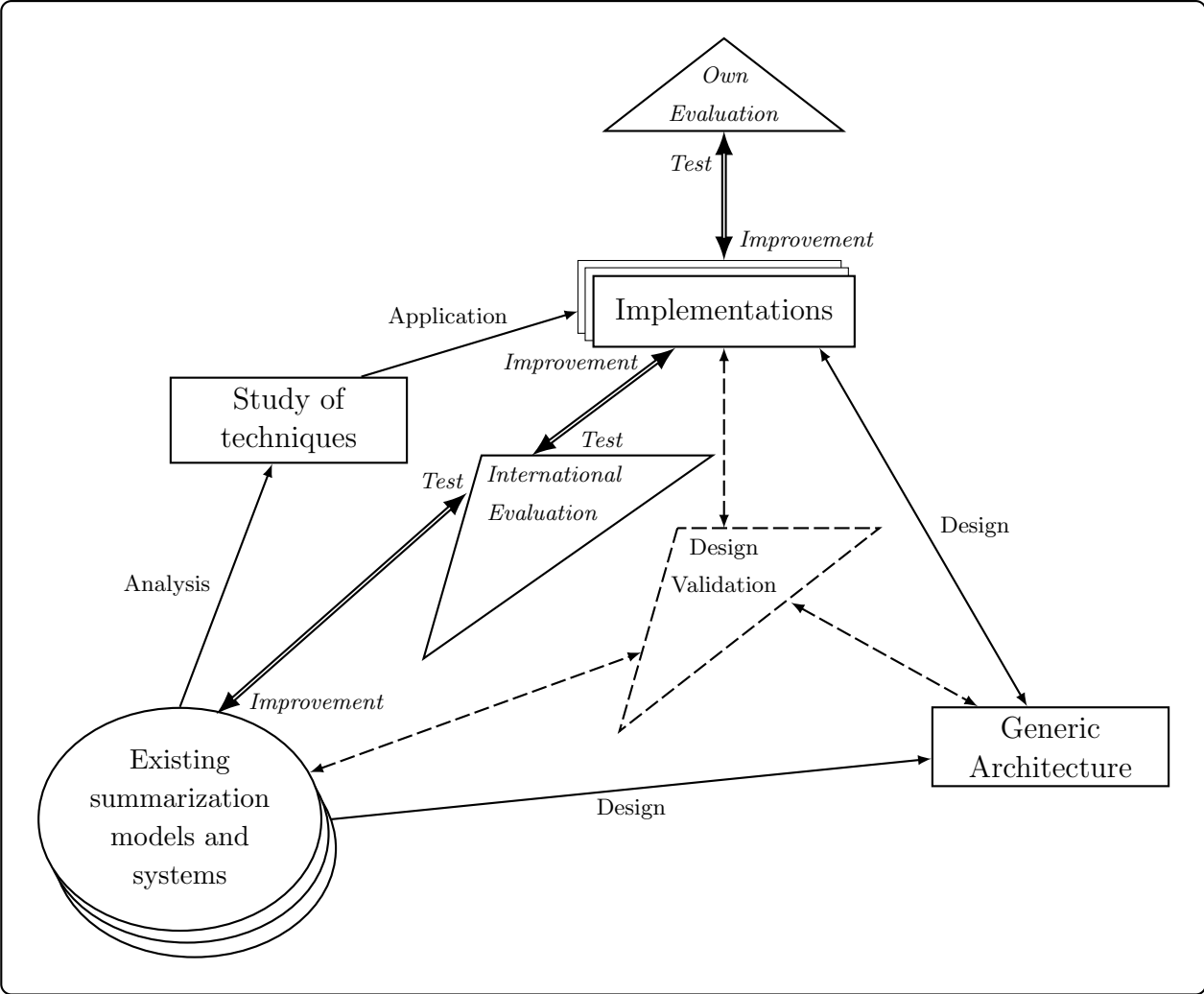
Figure 1.1: Diagram of the research method applied to obtain this thesis.

- Chapter 6 reports the performance obtained by several evolving prototypes in the query-focused Multi-Document Summarization task at the DUC international conference.

- Chapter 7 refers an experiment directed at analyzing the general architecture customization to summarize different sort of documents.

- Chapter 8 summarizes the author's thesis work with some conclusions and future directions in which the presented work could be extended.

To facilitate the reading process, a list of acronyms and other abbreviations used in this document is provided, as well as an index pointing to relevant words and concepts.

As it will be indicated, parts of the thesis material have been published.