FLEXIBLE SUMMARIZATION

J. F. Delannoy[1], K. Barker[2], T. Copeck[2], M. Laplante[1], S. Matwin[2], S. Szpakowicz[2]

[1] RES International

[2] School of Information Technology and Engineering, University of Ottawa

{delannoy, laplante}@res.ca, {kbarker, terry}@csi.uottawa.ca, {stan, szpak}@site.uottawa.ca

Abstract

Our project, initiated in 1997, approaches text summarization as a knowledge-scant task of passage selection. Several features make this task more discriminating. These features include "smart" key phrase selection that uses machine learning techniques and simple linguistic criteria; dynamic passage selection; adaptation to the type of text; and choice among several styles of summary. This paper presents the guiding principles of the project, describes the current state of the prototype, and discusses short-term and long-term future research.

1. Overview

1.1 Objectives

The text summarization project, part of work on intelligent information access, has been a joint effort of the Knowledge Acquisition and Machine Learning lab at the University of Ottawa, the National Research Council, and RES International. We have defined a model of a summarizer that tries to be realistic and robust, while introducing original features in all stages of processing.

- *Knowledge-scant approach.* Our goal is a system that must perform acceptably in a wide range of domains. This means almost no knowledge-intensive linguistic processing of texts. On the other hand, we do not want to forgo linguistic facts entirely in favour of purely statistical analyses. That is, summarization will rely on key phrases (identified in several ways) and on simple surface-syntactic indicators that can be interpreted with tolerable ambiguity.
- *Keyword/key phrase extraction* is done in two phases: the Extractor (Turney 1996), a system based on machine learning techniques, is used to pre-select text segments, which then undergo syntactically informed processing.
- *Composite rating of passages* will apply several criteria, including:
  - statistical distribution of exact and approximate *matches* of keywords and key phrases;
  - use of simple *discourse clues* (Berri 1996; Lehmam 1997; Minel *et al.* 1997), such as recognizing that "in summary" or "the bottom line" mark interesting passages;
  - domain-specific heuristics to determine the level of interest: linguistic (for example, identify head nouns in nouns phrases -- see below) and positional (the beginning of text, introduction or conclusion, beginning or end of paragraphs).
- *Passage selection with dynamic rating* will allow a coverage bias; variety in content is introduced by favouring sentences with keywords not yet represented in the selected set.
- The user can *parametrize* the process by constraining the *length* of the summary (absolute or relative to the size of the text). We plan to work on the parametrization (user-assisted and eventually automatic) of the *type of text* (news reports, user manuals...), and on the *type of summary* (for example generic or query-based; indicative or informative).
- *Multilingual capability* (currently, English, French, and Spanish): automatic detection of the language of the input.

The current prototype, written in Perl, implements part of these capabilities (see the Appendix).

To sum up: we will apply to summarization a combination of linguistic facts present at the text's surface and facts that can be computed with little effort. We will employ elements of computational linguistics and machine learning.

We have quoted in this paper work that we find most relevant to this paper, but there is no separate "Related work" section. A major recent collection of papers on text summarization is Mani and Maybury (1997).

1.2 Sources of information for summaries

The general idea of basing summarization on key phrases with a frequency above a certain threshold has been tested in a prototype that does not use any linguistic information yet. The following main sources of information are used:

- *Keywords and key phrases*: classically, the count or density of keywords is an essential component of the relevance score of passages.
- *Elements of text structure* include formatting (for example, numbered or bullet lists; HTML tags), and cues indicating articulation (for example, "first", "secondly", "finally" at the beginning of a clause; "in contrast with", "for ezample", etc.).
- *Surface indicators of text function*, especially definitions (as in Berri 1996).
- *Surface indicators of argument articulation* (certain standard phrases or function words) are taken into account. For example, a sentence beginning with "Thus" or "This shows that" is more likely to have a conclusion than a preceding sentence.

2. The current system

2.1 Organization of the system

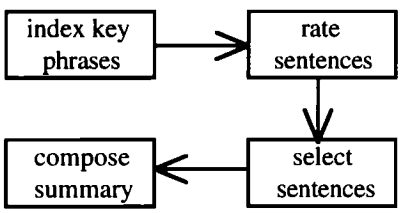

Figure 1: Processing stages

The processing stages (Fig. 1) are explained in the following sections.

2.2 Representation

We use an object-oriented representation for keywords and for sentences. A keyword object contains its string, and a flag used while assembling the summary to indicate whether the keyword has been found yet in any of the selected sentences.

A sentence object contains the original string, the trimmed string (after removing carriage returns, HTML code and so on), and a variety of ratings: for keyword count, for the presence of surface clues (for example, words like "important", "main" can be considered to indicate especially relevant passages), for positional information (is the sentence in the first paragraph; is it at the beginning of its paragraph).

All these criteria can be easily computed using pattern-matching and simple rules. This should make our system respond realistically fast, probably faster than a knowledge-intensive and linguistic-intensive approach (for instance, with full parsing).

The text is segmented into sentences. The algorithm to do this relies on several clues, including punctuation (for example, full stops when the last word is not on a known list of abbreviations), blank lines, and HTML formatting.

2.3 Key phrase selection

2.3.1 Language detection

The language of the input text is detected automatically, with promising reliability. (The user can also indicate a language explicitly). This is done by tallying the count of a dozen language-specific clues for each language considered, and choosing the best score. The clues are frequent words that do not exist in the other languages, for example, "cela, puis, avant, une" etc. for French. The (currently tentative) list of stop words corresponding to the language is read from a file.

2.3.2 Two methods of key phrase extraction

We are currently comparing the performance for summarization of the Extractor and a simple indexer that selects the phrases with the highest frequencies. The latter indexes the longest chains under constraints (no or few stop words) and their subchains. For example, seeing "mock turtle soup" leads to the indexing of *mock*, *mock turtle*, *mock turtle soup*, *turtle*, *turtle soup*, and *soup*.

A detailed evaluation of the Extractor against three other algorithms can be found in Turney (1997).

Indexed phrases with a single occurrence are dropped, since they cannot be used to evaluate the relevance of other sentences. The criterion for "best" combines the number of occurrences and the length of the key phrase.

2.4 Sentence rating

Sentences are initially rated by a linear combination of the values of the sentence descriptors.

Table 1 is a simplified display of sentence ratings, with the criteria used by the prototype so far: length, key phrase count, position in paragraph, and clue count.

Pronoun reference resolution, along with synonym expansion (using a thesaurus) and acronym expansion, if applied, must be done at this stage (this is not implemented yet). "Normalization" of synonyms to one preferred form would require morphological analysis, word-sense disambiguation, and a reliable thesaurus. This would also cover such phenomena as nominalization (for example, verb to noun, such as "decide" to "decision"). We will explore the feasibility, reliability and computing cost of this functionality.

2.5 Sentence selection

Sentences are selected on the merit of a global rating, calculated from the partial ratings attached to them via a profile (weight table) that corresponds to the text type and summary type. As a first approximation, we used this simple default profile:

```
score = kwdCnt + 1stInPar + clueCnt -
Pronouns
```

id	length	kwdCnt	1stInPar	clueCnt	string
1	63	1	1	0	http://miles.netpoint.net/~cubanet/CNews/y97/jan97/9albri.html
2	66	3	1	0	Clinton Extends His Suspension of Anti-Castro Law Another 6 Months
3	15	0	1	0	January 4, 1997
4	145	5	1	0	Caribbean: President says allies' own pressure on Cuba justifies postponing permission for Americans to sue foreign firms doing business there.
5	38	0	1	0	By STANLEY MEISLER, Times Staff Writer
6	285	9	1	0	WASHINGTON--President Clinton, citing evidence that "the international community is more united behind the cause of freedom in Cuba than ever before," ordered a second six-month suspension Friday of a law allowing American lawsuits against foreign companies doing business in Cuba.
7	135	3	1	0	The suspension further delays the implementation of the most contentious provision of a 1996 law that has infuriated American allies.
8	210	3	0	0	This provision of what is known as the Helms-Burton Act allows American citizens to sue foreign companies in U.S. courts if these companies are operating in Cuba on property originally owned by the Americans.
9	122	1	1	0	Both Clinton, who issued the order from his vacation site in the Virgin Islands, and Undersecretary of Commerce Stuart E.
10	170	4	1	0	Eizenstat, briefing reporters at the State Department, insisted that U.S. allies have now moved close enough to the American position on Cuba to justify the suspension.
11	257	4	1	0	But Eizenstat, who has traveled 50,000 miles in the last four months trying to assuage the feelings of the allies, acknowledged that Europe, Canada and Mexico all have retaliatory legislation in place should Clinton ever allow the lawsuits to go forward.
12	105	2	1	0	The presidential decision failed to fully satisfy either the originators of the law or its main critics.
13	221	2	1	0	In Washington, Sen. Jesse Helms (R-N.C.), chairman of the Senate Foreign Relations Committee and co-author of the act with Rep. Dan Burton (R-Ind.), said he was "extremely disappointed" by Clinton's "terrible mistake."
14	132	2	1	0	Helms acknowledged that "pressure from the Helms-Burton law has forced our European allies to take a tougher stand against Castro."

Table 1. An example of sentence ratings

That is: the initial score of a sentence again: is the sum of its keyword count, the "first in paragraph" parameter, and the count for the set of default contents and discourse clues, minus the count of pronouns.

We use an iterative selection method with a *coverage algorithm* relative to keywords. When a sentence is selected, we increase the score of all sentences that contain keywords still not included in the summary. (In every round, sentence ratings are incremented by the value of kwdCnt as recalculated over unused keywords.) This favours the inclusion in the summary of sentences mentioning all the keywords, and therefore, we hope, the concepts they denote. It also helps avoid redundancy, if we can assume that sentences containing the same keywords are more likely to have the same topic, or even the same contents.

Pronoun resolution has not been implemented yet. We will look at simple ways (recency; singular versus plural, or filtering of candidate referents by features like 'human' if available from a thesaurus) but they will not be fully satisfactory. Even with a reasonably reliable pronoun resolution algorithm, for hard cases, it may be useful to favour the selection of one or several sentences just before the sentence containing the pronoun, if the summary length restrictions allow it.

2.6 Summary production

The summary is eventually displayed alone or as highlighted passages in the original text. If HTML has been requested, keywords will be highlighted in colours. Connector words, pronouns, and stock phrases marking argumentation can be made to stand out, but so far they are not processed.

Table 2 shows an example summary, presented as highlighted passages in the text.

2.7 User interface

The interface (Fig. 2) allows the user to choose:

- the input file with a text for summarizing;
- the source of keywords: either the Extractor or the system's simple indexer.
- the language of the input (but the system can determine it on its own);
- output format: HTML or plain text;
- relative or absolute size of the summary;
- whether to display a trace, and tables showing the components of sentence rating;
- whether to display the summary alone, or as highlights in the input text, or both.

3. Future research

3.1 General directions

We will continue our research in the different directions we have already initiated.

- Uses of linguistic processing:

KEY PHRASES (auto) : cuba, clinton, law, foreign, president, suspension, business, companies, allies, eizenstat, international, democracy, mexican

http://miles.netpoint.net/~cubanet/CNews/y97/jan97/9albri.html

Clinton Extends His Suspension of Anti-Castro Law Another 6 Months

January 4, 1997

*(1) Caribbean: **President** says **allies'** own pressure on **Cuba** justifies postponing permission for Americans to sue **foreign** firms doing **business** there.*

By STANLEY MEISLER, Times Staff Writer

WASHINGTON--*(2) **President** Clinton, citing evidence that "the **international** community is more united behind the cause of freedom in **Cuba** than ever before," ordered a second six-month **suspension** Friday of a **law** allowing American lawsuits against **foreign companies** doing **business** in Cuba.*

The suspension further delays the implementation of the most contentious provision of a 1996 law that has infuriated American allies. This provision of what is known as the Helms-Burton Act allows American citizens to sue foreign companies in U.S. courts if these companies are operating in Cuba on property originally owned by the Americans.

Both Clinton, who issued the order from his vacation site in the Virgin Islands, and Undersecretary of Commerce Stuart E.

*(3) **Eizenstat**, briefing reporters at the State Department, insisted that U.S. **allies** have now moved close enough to the American position on **Cuba** to justify the **suspension**.*

*(4) But **Eizenstat**, who has traveled 50,000 miles in the last four months trying to assuage the feelings of the **allies**, acknowledged that Europe, Canada and Mexico all have retaliatory legislation in place should **Clinton** ever allow the lawsuits to go forward.*

The presidential decision failed to fully satisfy either the originators of the law or its main critics.

In Washington, Sen. Jesse Helms (R-N.C.), chairman of the Senate Foreign Relations Committee and co-author of the act with Rep. Dan Burton (R-Ind.), said he was "extremely disappointed" by Clinton's "terrible mistake."

Helms acknowledged that "pressure from the Helms-Burton law has forced our European allies to take a tougher stand against Castro." But he said Clinton had issued "what is, for all practical purposes, an indefinite suspension."

Table 2. An example summary

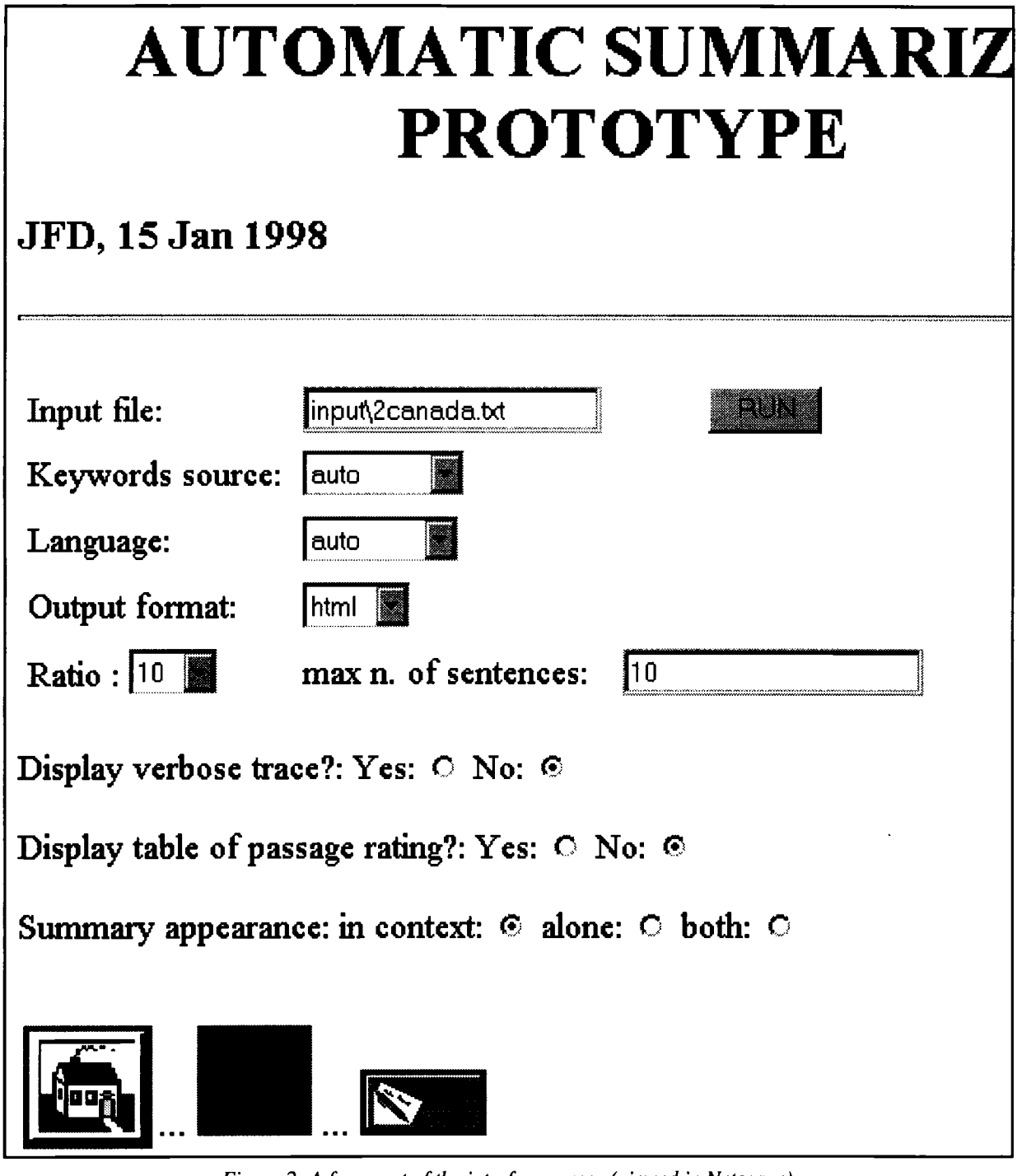

Figure 2: A fragment of the interface screen (viewed in Netscape)

- acquisition of keywords;
- approximate matching of key phrases; matching of acronyms and definitions.
- Word-sense disambiguation using WordNet—see for example Li *et al.* (1995).
- Recognition of simple rhetorical structures from superficial clues; see, for example, Ono *et al.* (1994).
- Sentence selection process: use of a thesaurus in matching with keywords.
- Simple assembly technique to avoid coherence gaps (caused by dangling referents and discourse connectors).

We develop here some ideas about the parametrization of the input and output.

3.2 Use of natural language processing techniques

3.2.1 Noun phrase seeker

We plan to detect noun phrases using a subset of DIPETT, a robust broad-coverage parser of English based on Quirk's grammar (Quirk *et al.* 1985); see Delisle 1994; Delisle & Szpakowicz 1995). We will evaluate the improvement brought about by considering authentic noun phrases as keywords, instead of cooccurrences (that is, word sequences, considered without tagging or parsing).

3.2.2 Key phrase selection from noun phrases

Noun phrase detection using the proposed DIPETT module will allow for key phrase extraction more advanced than that performed by the Extractor. Different surface representations of the same key phrase can be reconciled. For example, both "a processor's rated speed" and "the speed of the brand new (and aggressively advertised) Intel processors" could be considered to match the key phrase "processor speed". We find that head nouns in noun phrases make very good key phrases. On the other hand, the Extractor is faster and can extract key phrases without requiring external resources such as dictionaries.

A comparison of the two mechanisms for key phrase extraction shows that the linguistic approach produced key phrase lists with less noise (such as inclusion of more than one surface form of the same key phrase). Extractor, however, consistently produces acceptably accurate key phrase lists very quickly. As a result, for real time extraction from large texts, the Extractor is more suitable, and we propose to use its key phrases for preselection of sentences. The parser-based tool will be used on a smaller set of sentences to firm up the key phrase selection for final calculations.

Laurendeau (1992) studied a combination of heuristics to detect technical terms in a text. We will consider using such technical terms directly as keyphrases.

Another possible use of linguistic methods is the simplification of sentences (Lehmam 1997) by removing less interesting segments (such as paraphrases, examples), with recognition of syntactic boundaries.

3.3 Semi-automatic parametrization by category of texts

We did not conduct any significant evaluation so far, and still have to decide on the criteria to retain. Turney (1997) compares various keyword extractors to one another and to humans. For summarization, the task is more varied, and requires more attention to semantics. The goal is to compare the presence of "concepts", and maybe of argumentation. (Evaluation will be the theme of the first Automatic Text Summarization Conference (SUMMAC)—a call for participation has appeared very recently and cannot yet be properly addressed in this paper.)

It is interesting to bias summarization by type of text, that is, text purpose and style of presentation. Machine learning can be applied to determine which criteria "pay off" more in given situations; that is, it can set weights automatically so that for a certain type of text and certain summary type, the most discriminant types of parameters are given more importance. In fact, a key idea of the development of the Extractor was to learn the best descriptors; we plan to extend this principle to the sentence selection mechanism itself. Feedback would be given by human evaluators, or by comparison with preexisting "good" summaries.

3.3.1 Text style

There is a wide variety of text styles, and very different ways of presenting information. For example:

- first-page articles in newspapers: main facts early, then comments, background (mentioning again the names of the protagonists for their background, or the antecedents of the story as a background to the recent events);
- periodical/news-magazine: a long introduction, the real story in 3rd-5th paragraph; commentary often argumentative;
- scientific papers: the abstract and the conclusions usually discuss the contents explicitly.

This variety should be accounted for by different profiles, giving different sets of weights to the parameters describing the sentences. We have shown a simple set of weights in Section 2.7. We plan, with more testing and evaluation, to define a number of profiles corresponding to various text styles.

Profiles could also suggest different policies of handling references, and in particular pronouns. The use of pronouns varies widely with text style and purpose (one extreme is the style of patents, which tend to use no pronouns).

3.4. Examples of profiles

We will now show examples of qualitative and quantitative biases in profiles. [At the time of this writing the following descriptions are tentative. We are only beginning to design and test the profiles. In particular, some of the criteria for the clues rely on linguistic information that we do not yet handle in the prototype.]

3.4.1. Scientific paper

Heuristics

- strongly favour sentences with a high count of technical terms (TechTerms);
- position clues: first and last sentence in paragraph;

• content clues: sentence contains a definition; sentence contains "the/this paper article/report/paper".

Example profile:

```
score = kwdCnt + 2*TechTerms + contentClues + 1stInPar + lastInPar
```

Note: the content clues and discourse clues for this type of text are qualitatively different from those used for other types of input.

3.4.2. User manual for a consumer product

Heuristics

• content clues: the sentence follows a subtitle containing "how to", or it contains "how to" and has more than one clause;

• discourse clues: the sentence contains "important" or "caution" or "warning".

3.5 Summary type

The output category can be selected among several text categories.

- *Indicative summaries* tell what the text is about, for example, "This article describes a method for near-maximal yield electrolysis".
- *Informative summaries* give a concrete idea of what is said, for example "Electrolysis at near-maximal yield can be performed using wolfram-coated electrodes. We found an average gain in production of 1.5%, for equal power consumption".
- *Incremental or differential summaries* show how a text differs from others, in its contents or the way of presenting it, for example, for a set of news reports on the same event.
- A "*facts and figures*" summary gives a tabular synthesis of information, where this is justified or allowed by high homogeneity of information contents and style in one or across several sources. Jango (Etzioni 1997) does this for electronic catalogues on the WWW, filtering the input via customized "information adapters", rather than indexing and then choosing, it ignores whatever is not deemed relevant.

Alternatively, it would be interesting to study systematically existing and possible systems of annotation, from SGML to recent proposals of the Text Encoding Initiative (TEI), to make use of them in an innovative form of "summary" display. The annotations used could be very topical or contextual, according to its purpose.

3.6 How profiles could be used

The user could :

- choose one of several text types, or a neutral default;
- optionally, choose a "point of view" on the text's purposes, for example, to consider primarily facts or commentary;
- choose a summary type.

4. Conclusion

We are developing a composite summarizer which uses a variety of indices, either taken from the surface of the text and processed by a machine learning system (the Extractor), or resulting from fast linguistic processing assisted by parametrization by the user (to adapt to the input type and the desired style of the summary). The summarizer can also to some extent parametrize itself and learn from experience.

Much work remains before we achieve an efficient integration of our existing linguistic tools and make use of interesting discourse clues. At the same time the system must remain as fast and robust and we have set out in the definition of this project.

One of the main challenges is to fine-tune the heuristics that take accurately into account the style of the input and the desired output, bringing to bear an appropriate range of indices. In this task, machine learning methods could be used to select significant indices and optimize their weight. A possible bonus would be a system capable of text categorization over the same indices and the pattern in which the information is organized in the text.

Acknowledgement

This project is partially funded by the Natural Sciences and Engineering Research Council of Canada.

References

Barker, K., J.-F. Delannoy, S. Matwin and S. Szpakowicz, "Preliminary validation of a text summarization algorithm," University of Ottawa, Department of Computer Science, TR-96-04, 1996.

Berri, J. Cartier E., Desclés, J-P., Jackiewicz, A. Minel, J-L. (1996). Filtrage Automatique de textes , In *Natural Language Processing and Industrial Applications*, (pp 28-35), Université de Moncton, Nouveau-Brunswick, Canada

Copeck, T., S. Delisle and S. Szpakowicz, "Parsing and Case Analysis in TANKA," *15th Intl Conf on Computational Linguistics COLING-92*, Nantes, 1992, 1008-1012.

Delisle, S., *Text Processing without A-Priori Domain Knowledge: Semi-Automatic Linguistic Analysis for Incremental Knowledge Acquisition*, Department of Computer Science, University of Ottawa, PhD Thesis, 1994.

Delisle, S. and S. Szpakowicz, "Realistic Parsing: Practical Solutions of Difficult Problems," PACLING'95, Brisbane, 1995, 59-68.

Etzioni, Ö., "Moving up the Information Chain. Deploying Softbots on the World Wide Web", *AI Magazine*, Summer 1997, pp 11-18

Jacobs, P. S. and L. F. Rau, "Innovations in Text Interpretation," *Artificial Intelligence*, vol. 63(1-2), 1993, 143-191.

Laurendeau, C., "Automated Acquisition of Technical Concepts from Unrestricted English Text Using Noun Phrase Classification," Department of Computer Science, University of Ottawa, Master's thesis, 1992.

Lehmam A., "Automatic summarization on the Web? RAFI: a system for summarizing using indicating fragments", proceedings of *RIAO 97: Computer-assisted information searching on the Internet*, Montreal, June 1997, pp 112-123

Li, X., S. Matwin and S. Szpakowicz, "A WordNet-based Algorithm for Word Sense Disambiguation," *IJCAI-95*, Montreal, 1995, 1368-1374.

Mahesh, K.: Description of Analyst's Workstation and the HyperGen summarizer, Computing Research Laboratory, New Mexico State University, http://crl.nmsu.edu/users/mahesh/analyst/, http://crl.nmsu.edu/users/mahesh/summarization/

Mani, I. and M. Maybury, "Intelligent Scalable Text Summarization," *Proc ACL Workshop*, Madrid, 11 July 1997.

Mann, W. C. and S. A. Thompson, "Rhetorical Structure Theory: Toward a Functional Theory of Text Organization," *Text*, vol. 8(3), 1988, 243-281.

Matwin, S. and S. Szpakowicz, "Machine Learning Techniques in Knowledge Acquisition from Text," *THINK*, vol. 1(2), 1993a, 37-50.

Matwin, S. and S. Szpakowicz, "Text Analysis: How Can Machine Learning Help?," *First Conference of the Pacific Association for Computational Linguistics (PACLING)*, Vancouver, 1993b, 33-42.

Miike, S., E. Itoh, K. Ono and K. Sumita, "A Full-Text Retrieval System with a Dynamic Abstract Generation Function," *SIGIR*, 1994, 152-161.

Minel J.-L., S. Nugier et G. Piat: "Comment Apprécier la Qualité des "Résumés" Automatiques de Textes ? Les Exemples des Protocoles FAN et MLUCE et leurs Résultats sur SERAPHIN" , *JST Avignon* Avril 1997

Ono, K., K. Sumita and S. Miike, "Abstract Generation Based on Rhetorical Structure Extraction," COLING-94, 1994, 1-5.

Quirk, R., S. Greenbaum, G. Leech and J. Svartvik, *A Comprehensive Grammar of the English Language.* London: Longman, 1985.

Rino, L. and D. Scott, "Content Selection in Summary Generation," *XIth Brazilian Symposium on Artificial Intelligence (SBIA)*, Fortaleza, 1994, 411-423.

Salton, G., J. A., C. Buckley and A. Singhai, "Automatic Analysis, Theme Generation, and Summarization of Machine-Readable Texts," *Science*, vol. 264, 1994, 1421-1426.

Turney, P: "Selection of Key phrases from Text. Evaluation of Four Algorithms", NRC Technical Report ERB-1051 .

Turney, P , 1997:: downloadable site of the Extractor at http://ai.iit.nrc.ca/II_public/extractor

Note: more references used in the definition of our work objectives can be found in Barker et al. 1996, and at http://www.site.uottawa.ca/tanka/ts.html .

Appendix A: Functionalities of the current prototype (January 1998)

A.1 What the prototype can do

- handles texts in ASCII and HTML;
- uses the Extractor or its own method to retrieve key phrases (simple search for chains with no stopwords; indexing those chains and their subchains);
- allows the user to specify the summary length by ratio or maximum size;
- rates sentences by key phrase count, first-in-paragraph flag, discourse;
- processes clue count, length, and optionally the presence of proper nouns;
- selects sentences with dynamic re-rating;
- recognizes and handles English, French, and Spanish;
- displays summary alone or as highlighted source;
- processes local or remote files (by URL)

A.2 What the prototype cannot do

- apply linguistics techniques (tagging, stemming, NP isolation, detection of head nouns);
- handle connectors and negations;
- apply heuristics to recognize focus, definitions, etc.;
- flag or process quoted passages differently;
- apply heuristics to find persons' names;
- allow the user to add or remove keywords;
- use a thesaurus;
- simplify/reword sentences during summary production;
- recognize discourse functions in passages (for example, introductive or conclusive);
- use profiles for text type;
- automatically determine text type;
- produce different styles of summaries.