Producing Intelligent Telegraphic Text Reduction to provide an Audio Scanning Service for the Blind

Gregory Grefenstette
Xerox Research Centre Europe
Grenoble Laboratory
38240 Meylan, France
Gregory.Grefenstette@xrce.xerox.com

Abstract

Text summarization is usually taken to mean producing a shorter version of an original document by retaining the most salient parts of the original text. Two approaches have been favored: selecting high content-bearing sentences influenced by positional constraints, and performing domain-dependent information extraction which fills a template from which a summary can be glossed. This paper presents a third approach to text reduction, producing shortened telegraphic versions of all input sentences. This approach shares the domain-independence advantage of the sentence selection and the one-pass advantage of information extraction template filling. We argue that this type of text reduction has some practical applications, and present the case of creating an audio scanner for the blind.

Introduction

To the layman, text summarization(Paice 1990) means producing a concise version of a document highlighting both the important and original points of the documents. This idea of summarization readily accepts the use of words not found in the documents, admitting the generalization of key points (Sparck Jones 1996) which may only appear implicitly, or be mentioned in passing, in the original document. Such a task depends upon a number of resources that machine-based summarization systems generally do not have access to: (i) a model of what the user already knows, (ii) a model of the domain that could be used to paraphrase using a metalanguage, and (iii) a model of the domain that could be used to distinguish what is original.

In the summarization efforts arising from work in the information extraction community (DeJong 1982; Jacobs & Rau 1990; Robert Gaizauskas 1997) and in (Paice & Jones 1993), a limited domain model *does* exist, expressed for example as an event template such as a *company merger*. The template is often filled by recognizing arguments extracted by passing a number of lexical syntactic pattern-matchers over the input text. Once a template is filled, a gloss of its contexts, that is a natural language-like rewriting of the field values, can be produced as a shorter summary of the longer text. One of the advantages of this approach is that during the first pass over the text the summary can be produced in a stream mode, i.e. output is produced before the entire input is read. On the other hand, the major drawbacks to this approach are the initial costs of building up the event template and creating the lexical syntactic pattern extractors, and the system's subsequent rigidity, producing only a summary of the events for which it has been programmed.

From the computational linguistics community, a more open-ended and more widely applicable approach to summarization has been explored. This approach, exemplified by work from Rath *et al.* (Rath, Resnick, & Savage 1961) to Kupiec *et al.* (Kupiec, Pedersen, & Chen 1995), uses a domain-independent analysis of both document structure (where people put the most important information) and document terminology (what terms occur most frequently in a document) in order to rank all the sentences in the document from most important to least. Once the sentences are ranked, the user can specify the length of the summary desired, and the top N sentences can be returned in their original order as the summary.

Though this latter approach is by far the most

open and easily implementable approach to produce summaries for a wide variety of documents, there are a few cases in which this sentence extracting summarization is less adequate. For one, this type of summarization supposes that a document is about one thing and the summary should be geared toward that thing. The proposed techniques use the frequency of a term to decide its importance, so that if a composite document is being treated, one subject that dominates will become the principal subject reported, while other subtopics may be skipped. The extreme case of this would be when an important topic is mentioned only in passing, for example in a sentence not containing any of the highly weighted terms.

Another problem in some cases, for example when the document is being scanned bit by bit, is that the sentence-extraction technique requires a first pass over the entire document, since it uses the statistics on overall term frequency as part of its ranking criteria, before any output is produced. This makes it less adapted to the task of producing a summary during stream-based document capture.

A third potential problem, for short texts, also arises from this reliance on statistics to identify content: since frequency is used as an indicator of importance, words have to be repeated a certain number of times in order to be considered. In a shorter text, word counts are lower and are thus less reliable estimators for statistical methods.

In this paper, we present a third type of text reduction which, while not able to paraphrase events from specific domains as the information extraction techniques can, nor able to succinctly capture the essence of a long document as the sentence extraction techniques do, nonetheless reduces the size of the text while retaining much of the meaning and which has some practical applications for which it is well suited.

Scenario

Consider the following situation. Sighted readers can quickly scan over a page and understand the topic begin discussed. Blind readers cannot so easily jump from point to point in a text, and usually rely on speeding up and slowing down the speech production to perform the same function. If each text segment were summarized, the blind reader could jump from summary to summary in order to perform scanning. In addition to the two types of summarization techniques already proposed (template filling followed by glossing, or sentence extraction), we describe in the next section a third text reduction alternative, that we have implemented over a shallow parser (Grefenstette 1996a). This implementation provides a graduated sentence reduction similar to that proposed by Sager(Sager 1981)[p. 253-255] in her excision analysis application to language teaching.

If this system were integrated in a reading machine and these levels of sentence reduction were attached to a knob, a blind reader could skim the page in a way similar to sighted readers, processing only some of the meaningful parts of the input text. Telegraphic text reduction might be particularly adapted to the common situation that the blind reader faces when she is scanning in a document one page at a time, in a stream input mode. As the OCR is being performed on the document, telegraphic speech could be generated giving an idea of the content of the page being scanned.

A similar application could be useful for sighted people. Imagine audio scanning of web pages while driving a car. Telegraphic text could be synthesized at one level until some topic of interest appears, at which point the driver could give a vocal order for more detail on the current text. The current text would then be synthesized completely.

Telegraphic Text Reduction

Telegraphic text reduction works on the sentence level rather than the document level. Instead of identifying individual sentences as being important than other sentences, the telegraphic summarization technique identifies individual parts of each sentence as being more important than other parts of the same sentence. Of course, every element in a sentence carries some meaning, just as every sentence in a document adds to the sense of the document to be summarized. But some parts of a sentence can be removed, with less damage to the meaning, just as some sentences can be re-

moved from a complete document. For example, removing articles influences the meaning of a sentence but rarely destroys the idea of the topic[1]. We propose to do this reduction in a principled manner. The general principles, based on general linguistic intuitions and some results from information retrieval experiments(Jing & Croft 1994), can be summarized as: proper names are generally more important that common nouns, nouns are more important that adjectives, adjectives are more important that articles, nouns are more important than adverbs, negations are always important, subclauses are less important than main clauses, etc. That is, the judgment of importance in this method is based upon linguistic criteria rather than frequency information or rather than matching some predefined lexical item.

The technique works by marking up the input text with linguistic structural annotations. The 'intelligence' in our system comes from this linguistic mark-up that allows for a principled text reduction. These identified structures are then used to decide which words should be output according to the level of reduction requested. One particularity of our approach is that the all of the markup[2] and the subsequent text selection is made using finite-state techniques(Karttunen *et al.* 1996).

The first level of markup involves separating the input stream into tokens and sentences(Grefenstette & Tapanainen 1994). Each token undergoes a finite-state morphological analysis(Karttunen, Kaplan, & Zaenen 1992), at which point the token becomes annotated with all possible grammatical tags. This step is followed by part-of-speech disambiguation [3] after which only one part-of-speech tags remains appended to each token.

Using a technique similar to chunking (Debili 1982; Abney 1991), the part-of-speech annotated text is passed through a verb group annotator which inserts verb group markers around sequences of tagged words corresponding to a verb group.

Each verb group is passed through a finite-state filter(Grefenstette 1996a) that identifies the principal, or head, verbs of the verb group and a verb head marker is attached to each of these verbs. This verb head marker depends on the structure of the verb group. Example of verb group head markers are 'Head of Passive Verb Group', 'Head of Active Verb Group', 'Head of Infinitive Verb Group', etc.

In a similar way, the part-of-speech and verb group annotated text is passed through a noun group annotator which inserts noun group markers around sequences of tagged words corresponding to a noun group. Each noun group is passed through a finite-state filter which identifies the principal, or head, nouns of the noun group and a noun head marker is attached to each of these nouns. Examples of noun group head markers are 'FreeNoun' (which can be attached to a verb as a subject or object) and 'PrepNoun' which is attached to a preposition.

At this point, an additional set of filters (Grefenstette 1996a) identifies syntactic dependency relations, such as subject–verb, verb–object, verb–indirect objects, etc. A number of alternative methods for producing such relational markers have been proposed (Hindle 1993; Voutilainen, Heikkila, & Anttila 1992). The system used for the examples in this paper has about a 77% recall rate for dependency relations(Grefenstette 1996b) involving verbs and arguments. A more advanced system begin developed(Ait-Mokhtar & Chanod 1997) which has already attained a 93% average accuracy of subject recognition in newspaper text. The group and head annotated text is fed into the reducing transducer which eliminates words depending upon the depth of reduction specified by the user (or if none is specified a default depth can be used).

Some examples of levels of reductions, though other combinations could be imagined, are given in the following list. In all cases, negation must be retained once verbs are involved:

[1] For example, removing articles influences meaning of sentence but rarely destroys idea of topic.

[2] Only the part-of-speech disambiguation(Cutting *et al.* 1992) involves a treatment outside the realm of finite state technology.

[3] WWW demos for part-of-speech taggers are at http://www.xrce.xerox.com/research/mltt/Tools/pos.html

1. only proper names, no subclauses
2. only subjects and object nouns, no subclauses
3. only subjects, head verbs, and object nouns, no subclauses
4. only subjects, head verbs, and object nouns, preposition and dependent noun heads, no subclauses
5. only proper names, including subclauses
6. only subjects and object nouns, including subclauses
7. only subjects, head verbs, and object nouns, including subclauses
8. only subjects, head verbs, and object nouns, preposition and dependent noun heads, including subclauses

The reduction levels given above have the general effect of reducing the length of the text, retaining more or less words around the skeletal parts of the sentence.

The output is a telegraphic version of the text which can then be fed into a speech synthesizer. The last sample reduction on the list above is somewhat similar to eliminating all the stopwords from a text.

Example

In Figures 1 and 2, we present an example of the behavior of the various settings on a text drawn from a WWW-based newspaper. The original passage is:

> *Former Democratic National Committee finance director Richard Sullivan faced more pointed questioning from Republicans during his second day on the witness stand in the Senate's fund-raising investigation. While he grew flustered at times, the 33-year-old Sullivan stayed tight-lipped, downplaying concerns about former fund-raiser John Huang.*

This text is fed into our telegraphic reduction system, and one of the first steps is to mark up the text with the additional grammatical and syntactic markings shown in Figure 1.

Setting the reduction level at the levels described in the preceding sections produces one of the telegraphic versions of the input text shown in Figure 2. As an anecdotal illustration of the robustness of the reduction technique, Note that in Figure 1 that the adverb *more* is incorrectly retained in the verbal chain by the shallow parser, but that this does not influence the output presented in Figure 2 since the adverbs are removed in the levels presented. Similarly correct decisions about legitimate linguistic concerns such as the proper attachment of prepositional phrases do not influence the output of the reducer since either all prepositional phrases are kept or deleted. Current recognition of the limits of subclauses is still a problem.

Using a text-to-speech synthesizer such as the public-domain synthesizer rsynth[4], the original text is pronounced in 36 seconds. Pronouncing the text produced by the reduction labeled as level 8 reduction shown in Figure 2 (in which much of the original meaning is retained) takes only 20 seconds with the same text-to-speech synthesizer. For about a page-full of text, about 500 words or 11 times the text shown, the system performs all the steps from tokenization to tagging to markup to reduction and output in 5 seconds on a Sparc Ultra 1, much faster than the time needed to synthesize and pronounce the final speech. The appendix shows further examples. An online demonstration will soon be available on our web site http:// www.xrce.xerox.com/ research/mltt/toolhome.html.

Conclusion

We have presented an intelligent telegraphic text reduction method. The purpose of the method is to reduce the size of the text by eliminating linguistically identified parts of input sentences. Although telegraphic text production has not been one of the main axes of research in text summarization, it shares some of the advantages of the wide applicability of sentence selection techniques since it is not domain dependent, as well as some of the advantages of template filling since it can produce output in a stream mode. These two qualities make it well adapted to certain speech production tasks such as the one described in the scenario of the blind reader scanning pages one at a time.

[4] http://wwwtios.cs.utwente.nl/say/about_say.html ftp://svr-ftp.eng.cam.ac.uk/pub/comp.speech/synthesis/

This technique requires further, user-based research in order to see what level of reduction is most appropriate as a default, i.e. where intelligibility disappears.

References

Abney, S. 1991. Parsing by chunks. In Robert Berwick, S. A., and Tenny, C., eds., *Principle-Based Parsing.* Dordrecht: Kluwer Academic Publishers.

Ait-Mokhtar, S., and Chanod, J.-P. 1997. Incremental finite-state parsing. In *ANLP'97*, 72–79.

Cutting, D.; Kupiec, J.; Pedersen, J.; and Sibun, P. 1992. A practical part-of-speech tagger. *ANLP'92.*

Debili, F. 1982. *Analyse Syntaxico-Semantique Fondee sur une Acquisition Automatique de Relations Lexicales-Semantiques.* Ph.D. Dissertation, University of Paris XI, France.

DeJong, G. F. 1982. An overview of the FRUMP system. In Lehnert, W., and Ringle, M., eds., *Strategies for Natural Language Processing.* Hillsdale, NJ: Lawrence Erlbaum. 149–176.

Grefenstette, G., and Tapanainen, P. 1994. What is a word, what is a sentence? Problems of tokenization. In *3rd Conference on Computational Lexicography and Text Research.* Budapest, Hungary: COMPLEX'94. http://www.xrce.xerox.com/publis/mltt/mltt-004.ps.

Grefenstette, G. 1996a. Light parsing as finite state filtering. In *Workshop on Extended finite state models of language.* Budapest, Hungary: ECAI'96.

Grefenstette, G. 1996b. Using the SEXTANT low-level parser to analyse a software manual corpus. In Sutcliffe, R. F.; Koch, H.-D.; and Annette., eds., *Industrial Parsing of Software Manuals*, number 17 in Language and computers: Studies in Practical Linguistics. Amsterdam: Rodopi.

Hindle, D. 1993. A parser for text corpora. In Atkins, B., and Zampolli, A., eds., *Computational Approaches to the Lexicon.* Clarendon Press.

Jacobs, P., and Rau, L. 1990. SCISOR: Extracting information from on-line news. *Communications of the ACM* 33(11):88–97.

Jing, Y., and Croft, W. B. 1994. An Association Thesaurus for Information Retrieval. In *RIAO'94, Intelligent Multimedia Information Retrieval Systems and Management*, 285–298.

Karttunen, L.; Chanod, J.; Grefenstette, G.; and Schiller, A. 1996. Regular expression for language engineering. *Natural Language Engineering* 2(4).

Karttunen, L.; Kaplan, R. M.; and Zaenen, A. 1992. Two-level morphology with composition. In *Proceedings COLING '92*, 141–148.

Kupiec, J.; Pedersen, J.; and Chen, F. 1995. A trainable document summarizer. In *Proceedings of the 18th Annual International ACM/SIGIR Conference*, 68–73.

Paice, C. D., and Jones, P. A. 1993. The identification of important concepts in highly structured technical papers. In *Proceedings of the 19th Annual International ACM/SIGIR Conference*, 69–78.

Paice, C. 1990. Constructing literature abstracts by computer: Techniques and prospects. *Information Processing and Management* 26:171–186.

Rath, G. J.; Resnick, A.; and Savage, R. 1961. The formation of abstracts by the selection of sentences: Part 1: sentence selection by man and machines. *American Documentation* 12(2):139–141.

Robert Gaizauskas, A. M. R. 1997. Coupling information retrieval and informatin extraction: A new text technology for gathering information from the web. In *RIAO'97, Computer-Assisted Information Searching on the Internet*, 356–376.

Sager, N. 1981. *Natural Language Information Processing.* Reading, Mass.: Addison–Wesley.

Sparck Jones, K. 1996. Summarization. In Cole, R. A., ed., *Survey of the State of the Art in Human Language Technology.* National Science Foundation, European Commission. chapter 7. http://www.cse.ogi.edu/CSLU/HLTsurvey/.

Voutilainen, A.; Heikkila, J.; and Anttila, A. 1992. A lexicon and constraint grammar of english. In *Proceedings of the Fourteenth International Conference on Computational Linguistics.* Nantes, France: COLING'92.

Appendix 1:Dublin (newspaper 12.8.1990)

The following example is from the sample texts given in Karen Sparck Jones Dagstuhl presentation: *Summarising: analytic framework, key component, experimental method* available on the web at http://www.fh-hannover.de/ ik/ Dagstuhl/ Abstract/ Abstracts/ Sparck-Jones/ Sparck-Jones.html

Two bank robbers were shot dead last Friday when a high-speed car chase ended in a gun battle in the city centre. IRA involvement has not been ruled out by detectives. The drama begun shortly after 10am when the men held up a bank 5 miles west of Dublin. They made off with 2,500 pounds but their car was spotted by armed detectives on the edge of the city and a high-speed car chase followed. The Garda car came under repeated fire from the raiders. The chase ended when the robbers tried to smash through a second Garda car that was blocking the road. A prolonged shootout followed as more armed Gardai arrived. The two men inside were reported dead on arrival in hospital. Gardai could not say who fired the first shot.

The level 8 reduction of the above text gives the following:

robbers shot last chase ended in battle in centre. IRA involvement not ruled by detectives. drama begun 10am men held up bank miles of Dublin. They made off with pounds but car armed detectives on edge of city and chase followed. Garda car came under fire from raiders. chase ended robbers tried smash through Garda car blocking road. prolonged shootout followed armed Gardai arrived. men reported on arrival in hospital. Gardai could not say fired first shot.

Appendix 2: Our abstract

When the same method is applied to the abstract of this paper, we get:

Text summarization producing version of document by retaining parts of text. Two approaches favored sentences influenced by constraints and extraction fills template from summary glossed. This paper presents approach to reduction shortened versions of input sentences. This approach shares advantage of selection and advantage of filling. We argue type of reduction has applications and present case of creating scanner for blind.

The beginning of *Walden Pond*:

When I wrote the following pages, or rather the bulk of them, I lived alone, in the woods, a mile from any neighbor, in a house which I had built myself, on the shore of Walden Pond, in Concord, Massachusetts, and earned my living by the labor of my hands only. I lived there two years and two months. At present I am a sojourner in civilized life again. I should not obtrude my affairs so much on the notice of my readers if very particular inquiries had not been made by my townsmen concerning my mode of life, which some would call impertinent, though they do not appear to me at all impertinent, but, considering the circumstances, very natural and pertinent. Some have asked what I got to eat; if I did not feel lonesome; if I was not afraid; and the like. Others have been curious to learn what portion of my income I devoted to charitable purposes; and some, who have large families, how many poor children I maintained. I will therefore ask those of my readers who feel no particular interest in me to pardon me if I undertake to answer some of these questions in this book. In most books, the I, or first person, is omitted; in this it will be retained; that, in respect to egotism, is the main difference. We commonly do not remember that it is, after all, always the first person that is speaking. I should not talk so much about myself if there were anybody else whom I knew as well. Unfortunately, I am confined to this theme by the narrowness of my experience. Moreover, I, on my side, require of every writer, first or last, a simple and sincere account of his own life, and not merely what he has heard of other men's lives; some such account as he would send to his kindred from a distant land; for if he has lived sincerely, it must have been in a distant land to me. Perhaps these pages are more particularly addressed to poor students. As for the rest of my readers, they will accept such portions as apply to them. I trust that none will stretch the seams in putting on the coat, for it may do good service to him whom it fits.

I wrote pages or bulk of them I lived in woods mile from neighbor in house I built myself on shore of Walden Pond in Concord Massachusetts and earned living by labor of hands only . I lived there years and months . I am sojourner in life . I not obtrude affairs on notice of readers inquiries not made by townsmen concerning mode of life some call impertinent they not appear to me impertinent but considering circumstances natural and pertinent . Some asked I got eat I not feel lonesome I was not afraid and the like . Others been curious learn portion of income I devoted to purposes and some have families children I maintained . I ask those of readers feel no interest in me pardon me I answer some of questions in book . In books I or person omitted in this it retained that in respect to egotism is difference . We not remember it is person speaking . I should not talk myself there were anybody I knew . I confined to theme by narrowness of experience . Moreover I on side require of writer first or last simple and account of life and not he heard of men 's lives account he send to kindred from land for he lived it been in land to me . pages addressed to students . As for rest of readers they accept portions apply to them . I trust none stretch seams in putting on coat for it do service to him it fits .

```
[NC Former/JJ Democratic/JJ National/JJ Committee/NN finance/NN director/NN Richard/NP
*HeadN Sullivan/NP NC]  [VC *ActV faced/VBD more/AP VC]  [NC *FreeAdj pointed/JJ
*HeadN questioning/VBG from/IN *PrepN Republicans/NNS during/IN his/PP$
second/OD *PrepN day/NN on/IN the/AT witness/NN *PrepN stand/NN in/IN the/AT
*PrepN Senate/NP 's/POS *IngV fund-raising/VBG *PrepN investigation/NN NC] .

While/CS [NC *HeadN he/PPS NC] [VC grew/VBD *ActV flustered/VBN VC]
[NC at/IN *PrepN times/NNS NC] ,/CM [NC the/AT 33-year-old/CD *HeadN Sullivan/NP NC]
[VC *ActV stayed/VBD VC] [NC *HeadN tight-lipped/JJ NC] ,/CM downplaying/VBG
[NC *HeadN concerns/NNS about/IN former/JJ fund-raiser/JJ John/NP *PrepN Huang/NP NC]
```

Figure 1: Syntactically marked-up text after shallow parsing. Using the following markups:

`[NC ... NC]`	a nominal grouping
`[VC ... VC]`	a verbal grouping
`*HeadN`	a nominal head not attached to a preposition
`*PrepN`	a nominal head attached to a preposition
`*FreeAdj`	an adjective whose attachment is undetermined
`*ActV`	head of an active verb group
`*ActV`	head of a gerundive verb group
`/JJ /NN /NP ...`	part of speech tags

reduction level	output
level 1	*Richard Sullivan Republicans Senate Sullivan John Huang.*
level 2	*Richard Sullivan pointed questioning. Sullivan tight-lipped concerns.*
level 3	*Richard Sullivan faced pointed questioning. Sullivan stayed tight-lipped concerns.*
level 4	*Richard Sullivan faced pointed questioning from Republicans during day on stand in Senate fund-raising investigation. Sullivan stayed tight-lipped concerns about John Huang.*
level 5	*Richard Sullivan Republicans Senate Sullivan John Huang.*
level 6	*Richard Sullivan pointed questioning. Sullivan tight-lipped concerns.*
level 7	*Richard Sullivan faced pointed questioning from Republicans during day on stand in Senate fund-raising investigation. he flustered, Sullivan stayed tight-lipped concerns about John Huang.*
level 8	*Richard Sullivan faced pointed questioning from Republicans during day on stand in Senate fund-raising investigation. he flustered at times Sullivan stayed tight-lipped concerns about John Huang.*

Figure 2: Eight levels of text reduction.