

# Using Coreference Chains for Text Summarization

Saliha Azzam and Kevin Humphreys and Robert Gaizauskas  
{s.azzam,k.humphreys,r.gaizauskas}@dcs.shef.ac.uk

Department of Computer Science, University of Sheffield  
Regent Court, Portobello Road  
Sheffield S1 4DP UK

## Abstract

We describe the use of coreference chains for the production of text summaries, using a variety of criteria to select a ‘best’ chain to represent the main topic of a text. The approach has been implemented within an existing MUC coreference system, which constructs a full discourse model of texts, including information about changes of focus, which can be used in the selection of chains. Some preliminary experiments on the automatic evaluation of summaries are also described, using existing tools to attempt to replicate some of the recent SUMMAC manual evaluations.

## 1 Introduction

In this paper we report preliminary work which explores the use of coreference chains to construct text summaries. Sparck Jones (1993) has described summarization as a two stage process of (1) building a representation of the source text and (2) generating a summary representation from the source representation and producing an output text from this summary representation. Our source representation is a set of coreference chains – specifically those chains of referring expressions produced by an information extraction system designed to participate in the MUC-7 coreference task (DARPA, 1998). Our summary representation is a ‘best chain’, selected from the set of coreference chains by the application of one or more heuristics. The output summary is simply the concatenation of (some subset of) sentences from the source text which contain one or more expressions occurring in the selected coreference chain.

The intuition underlying this approach is that texts are in large measure ‘about’ some central entity, which is effectively the topic, or focus of the discourse. This intuition may be false –

there may be more than one entity of central concern, or events or relations between entities may be the principal topic of the text. However, it is at the very least an interesting experiment to see to what extent a principal coreference chain can be used to generate a summary. Further, this approach, which we have implemented and preliminarily evaluated, could easily be extended to allow summaries to be generated from (parts of) the best  $n$  coreference chains, or from event, as well as object, coreference chains.

The use of document extracts formed from coreference chains is not novel. Bagga and Baldwin (1998) describe a technique for cross-document coreference which involves extracting the set of all sentences containing expressions in a coreference chain for a specific entity (e.g. *John Smith*) from each of several documents. They then employ a thresholded vector space similarity measure between these document extracts to decide whether the documents are discussing the same entity (i.e. the same *John Smith*). Baldwin and Morton (1998) describe a query-sensitive (i.e. user-focused) summarization technique that involves extracting sentences from a document which contain phrases that corefer with expressions in the query. The resulting extract is used to support relevancy judgments with respect to the query.

The use of chains of related expressions in documents to select sentences for inclusion in a generic (i.e. non-user-focused) summary is also not novel. Barzilay and Elhadad (1997) describe a technique for text summarization based on *lexical* chains. Their technique, which builds on work of Morris and Hirst (1994), and ultimately Halliday and Hasan (1976) who stressed the role of lexical cohesion in text coherence, is to form chains of lexical items across a text based on the items’ semantic relatedness as in-

dictated by a thesaurus (WordNet in their case). These lexical chains serve as their source representation, from which a summary representation is produced using heuristics for choosing the ‘best’ lexical chains. From these the summary is produced by employing a further heuristic to select the ‘best’ sentences from each of the selected lexical chains.

The novelty in our work is to combine the idea of a document extract based on coreference chains with the idea of chains of related expressions serving to indicate sentences for inclusion in a generic summary (though we explore the use of coreference between query and text as a technique for generating user-focused summaries as well).

Returning to Halliday and Hasan, one can see how this idea has merit within their framework. They identify four principal mechanisms by which text coherence is achieved – reference, substitution and ellipsis, conjunction and lexical cohesion. If lexical cohesion is a useful relation to explore for getting at the ‘aboutness’ of a text, and hence for generating summaries, then so too may reference (separately, or in conjunction with lexical cohesion). Indeed, identifying chains of coreferential expressions in text has certain strengths over identifying chains of expressions related merely on lexical semantical grounds. For, there is no doubt that common reference, correctly identified, directly ties different parts of a text together – they are literally ‘about’ the same thing; lexical semantic relatedness, as indicated by an external resource, can never conclusively establish this degree of relatedness, nor indeed can the resource guarantee that semantic relatedness will be found when it exists. Further, lexical cohesion techniques ignore pronominal anaphora, and hence their frequency counts of key terms, used both for identifying best chains and best sentences within best chains, may often be inaccurate, as focal referents will often be pronominalised.

Of course there are drawbacks to a coreference-based approach. Lexical cohesion relations are relatively easy to compute and do not rely on full text processing – this makes summarisation techniques based on them rapid and robust. Coreference relations tend to require more complex techniques to compute. Our view, however, is that summarisation re-

search is still in early stages and that we need to explore many techniques to understand their strengths and weaknesses in terms of the type and quality of the summaries they produce. If coreference-based techniques can yield good summaries, this will provide impetus to make coreference technologies better and faster.

The basic coreference chain technique we describe in this paper yields *generic* summaries as opposed to *user-focused* summaries, as these terms have been used in relation to the TIPSTER SUMMAC text summarization evaluation exercise (Mani et al., 1998). That is, the summaries aim to satisfy a wide readership by supplying information about the ‘most important’ entity in the text. But of course this technique could also be used to generate summaries tailored to a user(group) through use with a preprocessor that analyzed a user-supplied topic description and selected one or more entities from the topic description to use in filtering coreference chains found in the full source document.

The rest of this paper is organised as follows. In Section 2 we briefly describe the system we use for computing coreference relations. Section 3 describes various heuristics we have implemented for extracting a ‘best’ coreference chain from the set of coreference chains computed for a text; and, it discusses how we select ‘best’ sentences to include in the summary from those source text sentences containing referring expressions in the ‘best’ chain. Section 4 presents a simple example and shows the different summaries that different heuristics produce. Section 5 describes the limited evaluation we have been able to carry out to date, but more importantly introduces what we believe to be a novel and interesting way of reusing some of the MUC materials for assessing summaries.

## 2 Coreference in the LaSIE system

The LaSIE system (Gaizauskas et al., 1995) has been designed as a general purpose IE system which can conform to the MUC task specifications for named entity identification, coreference resolution, IE template element and relation identification, and the construction of scenario-specific IE templates. The system has a pipeline architecture which processes a text one sentence at a time and consists of three prin-

cial processing stages: lexical preprocessing, parsing plus semantic interpretation, and discourse interpretation. The overall contributions of these stages may be briefly described as follows (see (Gaizauskas et al., 1995) for further details):

**lexical preprocessing** reads and tokenises the raw input text, performs phrasal matching against lists of proper names, identifies sentence boundaries, tags the tokens with parts-of-speech, performs morphological analysis;

**parsing and semantic interpretation**

builds lexical and phrasal chart edges in a feature-based formalism then does two pass chart parsing, pass one with a special named entity grammar, pass two with a general grammar, and, after selecting a 'best parse', which may have only partial coverage, constructs a predicate-argument representation of each sentence;

**discourse interpretation** adds the information from the predicate-argument representation to a hierarchically structured semantic net which encodes the system's world and domain model, adds additional information presupposed by the input, performs coreference resolution between new and existing instances in the world model, and adds any information consequent upon the new input.

The domain model is encoded as a hierarchy of domain-relevant concept nodes, each with an associated attribute-value structure describing properties of the concept. As a text is processed, instances of concepts mentioned in a text are added to the domain model, populating it to become a text-, or discourse-, specific model.

Coreference resolution is carried out by attempting to merge each newly added instance with instances already present in the discourse model. The basic mechanism, detailed in Gaizauskas and Humphreys (1997), is to examine, for each pair of newly added and existing instances: semantic type consistency/similarity in the concept hierarchy; attribute value consistency/similarity, and a set of heuristic rules, some specific to particular types of anaphora

such as pronouns, which can act to rule out a proposed merge. These rules can refer to various lexical, syntactic, semantic, and positional information about instances, and have mainly been developed through the analysis of training data. A recent addition, however, has been the integration of a more theoretically motivated focus-based algorithm for the resolution of pronominal anaphora (Azzam et al., 1998). This includes the maintenance of a set of focus registers within the discourse interpreter, to model changes of focus through a text and provide additional information for the selection of antecedents.

The discourse interpreter maintains an explicit representation of coreference chains created as a result of instances being merged in the discourse model. Each instance has an associated attribute recording its position in the original text in terms of character positions. When instances are merged, the result is a single instance with multiple positions which, taken together, represent a coreference chain.

### 3 Coreference Chain Selection

The summarisation mechanism is implemented as an additional module in the LaSIE system. It processes all the coreference chains built by the discourse interpreter and applies various criteria, described below, to select a 'best' chain. The set of sentences in which entries in the best chain occur are then identified, using their character positions, and concatenated together to act as a summary.

#### 3.1 Selection Criteria

The summarisation module implements several selection criteria which can be applied either independently or in combination, as specified in parameters of the module. The current set of criteria is not fixed, however, and is simply intended to represent common intuitive heuristics as a starting point for further experimentation.

**Length of Chain** This criteria simply prefers the chain containing the most entries, which represents the most frequently mentioned instance in a text, and so possibly the most important instance. In the case of several chains of equal maximum length, the spread is used in addition to select a single chain.

**Spread of Chain** This criteria involves the calculation of the distance, in byte offsets, between the earliest and latest entry in each chain. The chain which spans the greatest portion of the original text is preferred, corresponding to the intuition that an instance mentioned throughout a text, rather than just being frequently mentioned in one specific section, may be the most important instance. If this criteria fails to select a single chain, the length criteria will be used in addition.

**Start of Chain** A further intuition is that instances mentioned at the start of a text, or in a title if present, may be more important than those only introduced part way through. This criteria acts to require that the earliest entry in a chain is within either the title or the first paragraph of a text. Again, other criteria will be needed to select a single chain if several start in the first paragraph. This criteria may be more appropriate for particular text genres, such as newswires, than the previous two.

### 3.2 Focus Chains

An additional selection mechanism, which may be combined with the above criteria in several ways, is the 'reduction' of each coreference chain to what may be called a "focus chain". This makes use of the focus registers built within the discourse interpreter to track changes in the main focus of each clause in the text. A focus chain is the subset of a coreference chain which contains only those mentions of an instance that occur as the focus of a clause. For example, an indefinite noun phrase occurring as a logical subject may be recorded in the focus registers, and so be retained in a focus chain, whereas a subsequent definite noun phrase embedded in an optional prepositional phrase will not be in focus and will be removed.

The criteria listed above can all be applied to focus chains in exactly the same way as coreference chains. However, the point at which the coreference chains are reduced is significant: selecting the longest coreference chain and subsequently removing all non-focus entries before output may give different results than an initial selection of the longest focus chain. The former sequence of operations acts to filter the entries

within the 'best' coreference chain, possibly to reduce the final length of a summary, whereas the latter selects on the basis of the frequency (in combination with the length criteria) or the position (with spread) of focus itself. Only the latter sequence is considered in any detail below, although the implementation does allow experimentation with the use of focus chains at several alternative stages.

## 4 Example Output

This section illustrates the kind of summaries produced by the different heuristics, using an example from the MUC-6 evaluation corpus of newswire articles (much reduced for inclusion here by omitting the eight final paragraphs):

<DOC>

<HL> *Economy: Washington, an Exchange Ally, Seems To Be Strong Candidate to Head SEC*

</HL>

<TXT>

*Consuela Washington, a longtime House staffer and an expert in securities laws, is a leading candidate to be chairwoman of the Securities and Exchange Commission in the Clinton administration.*

*Ms. Washington, 44 years old, would be the first woman and the first black to head the five-member commission that oversees the securities markets.*

*Ms. Washington's candidacy is being championed by several powerful lawmakers including her boss, Chairman John Dingell (D., Mich.) of the House Energy and Commerce Committee. She currently is a counsel to the committee. Ms. Washington and Mr. Dingell have been considered allies of the securities exchanges, while banks and futures exchanges have often fought with them.*

*A graduate of Harvard Law School, Ms. Washington worked as a lawyer for the corporate finance division of the SEC in the late 1970s. She has been a congressional staffer since 1979.*

*Separately, Clinton transition officials said that Frank Newman, 50, vice chairman and chief financial officer of BankAmerica Corp., is expected to be nominated as assistant Treasury secretary for domestic finance. Mr. Newman, who would be giving up a job that pays \$1 million a year, would oversee the Treasury's auctions of government securities as well as banking issues. He would report directly to Treasury Secretary-designate Lloyd Bentsen.*

(...)

Using the 'length of chain' criteria selects the coreference chain for the person *Consuela*

Washington, and the following summary is produced:

<HL> Economy: Washington, an Exchange Ally, Seems To Be Strong Candidate to Head SEC  
</HL>

Consuela Washington, a longtime House staffer and an expert in securities laws, is a leading candidate to be chairwoman of the Securities and Exchange Commission in the Clinton administration.

Ms. Washington, 44 years old, would be the first woman and the first black to head the five-member commission that oversees the securities markets.

Ms. Washington's candidacy is being championed by several powerful lawmakers including her boss, Chairman John Dingell (D., Mich.) of the House Energy and Commerce Committee. She currently is a counsel to the committee. Ms. Washington and Mr. Dingell have been considered allies of the securities exchanges, while banks and futures exchanges have often fought with them.

A graduate of Harvard Law School, Ms. Washington worked as a lawyer for the corporate finance division of the SEC in the late 1970s. She has been a congressional staffer since 1979.

Using the 'spread of chain' selects the coreference chain for the person *Clinton*, used here as a noun modifier:

Consuela Washington, a longtime House staffer and an expert in securities laws, is a leading candidate to be chairwoman of the Securities and Exchange Commission in the Clinton administration.

Separately, Clinton transition officials said that Frank Newman, 50, vice chairman and chief financial officer of BankAmerica Corp., is expected to be nominated as assistant Treasury secretary for domestic finance.

Restricting the selection to focus chains only, the 'length of chain' criteria again chooses the *Ms. Washington* chain, but this does not include the elements occurring in phrases where *Ms. Washington* is not in focus. The summary is therefore reduced by omitting the non-focus mentions:

<HL> Economy: Washington, an Exchange Ally, Seems To Be Strong Candidate to Head SEC  
</HL>

Consuela Washington, a longtime House staffer and an expert in securities laws, is a leading candidate to be chairwoman of the Securities and Exchange Commission in the Clinton administration.

Ms. Washington, 44 years old, would be the first

woman and the first black to head the five-member commission that oversees the securities markets.

She currently is a counsel to the committee.

A graduate of Harvard Law School, Ms. Washington worked as a lawyer for the corporate finance division of the SEC in the late 1970s. She has been a congressional staffer since 1979.

This same summary is also obtained when the 'spread of chain' criteria is applied to the focus chains. The *Clinton* chain selected above does not include any focus elements, and the *Consuela Washington* focus chain is the most spread as well as the longest.

## 5 Evaluating Summaries

Evaluating the merit of a summary is a difficult task. The most extensive effort to date to develop a framework for assessing summaries has been the TIPSTER SUMMAC evaluation exercise (Mani et al., 1998). In this section we first review the SUMMAC evaluation measures and propose how we might 'simulate' these without expensive human judges; then, we describe how we have carried out one of these simulated evaluations to evaluate the summarization technique described in the previous sections.

### 5.1 The SUMMAC Evaluation

This exercise involved a number of different tasks and a number of different evaluation measures. The measures divided into *extrinsic* measures – those that ignore the content of the summary and assess it solely according to how useful it is in enabling an agent to perform some measurable task – and *intrinsic* measures – those that examine the content of the summary and attempt to pass some judgment on it directly. In brief, the four SUMMAC tasks were:

1. *Ad Hoc Task* The summariser is given a topic description and a set of documents and produces user-focused summaries based on the topic description. The summaries and the topic description (along with some source documents for control) are passed to a judge who reads the summaries passes relevance judgments on them with respect to the topic. These judgments are scored against 'true' relevance judgments previously established for the full documents with respect to the topic. This

is an extrinsic evaluation that measures the utility of a summarization system at generating user-focused summaries capable of supporting a relevance judgment.

2. *Categorization Task* The summariser is given a set of documents only and generates a summary of each. The summaries (again along with some source documents for control) along with five topic descriptions are given to a judge who reads the summaries and categorizes them with respect to the five topics, or “none of the above”, if none is deemed appropriate. These categorization judgments are scored against ‘true’ categorization judgments previously established for the full documents with respect to the topics. This is an extrinsic evaluation that measures the utility of a summarization system at generating generic summaries capable of supporting a categorization judgment.
3. *Question-Answering Task* The summariser is given a set of documents and a small set of topic descriptions (three in the actual evaluation) and produces ‘informative’, user-focused summaries. A judge is first given the topic descriptions and for each produces a set of topic questions, each capturing an ‘obligatory’ aspect of the topic – something that has to be present in a document to make it relevant to the topic. The judge then reads the set of relevant, full documents to be used in the evaluation and for each produces an answer key that contains answers to each of the topic questions. Finally the judge reads the summaries and scores each against the answer key for the document, deciding for each topic question whether the summary supplies a correct, partially correct, or missing answer. This is an intrinsic evaluation that measures the utility of a summarization system at generating ‘informative’ user-focused summaries capable of acting as surrogates for the original document for an agent charged with answering the questions implicit in a topic description.
4. *Acceptability Task* The summariser is given a set of documents only and generates a summary of each. The summaries and

the full documents from which they were generated are given to a judge who reads the summaries and documents and passes a binary ‘goodness’ judgment on the summaries, with no prior guidelines specifying what constitutes a good summary. This is an intrinsic evaluation that was carried out merely to see what sort of correlation might emerge between the extrinsic evaluations and naive intrinsic evaluation.

This is a very valuable framework, but for those without access to the resources to instantiate the framework, particularly the judging functions, it is hard to use it to assess a summarization system. Thus, we have considered ways to automate the judging function and have identified several, which while by no means perfect, may at least provide some useful information about the value of the summaries produced.

For the ad hoc task, we propose replacing the judge with an IR system that given the topic description assesses both the summaries and full documents for relevance. These judgments may then be compared against the ‘true’ relevance judgments. The IR system’s ability to judge full documents may be taken as a baseline measure, and the (predicted) loss in accuracy of that system relative to the baseline when assessing the summaries should provide a measure of how good the summaries are. This will not be an absolute measure, of course, but one which may serve to compare different summarization systems, or different parameter settings of the same system.

Similarly, for the categorization task, a document categorization system could be benchmarked against the full documents, and then, given this baseline, run against the output of the summarisation system. Again, an assessment of summaries relative to the baseline would be obtained which could provide significant insights, without the cost of a human judge.

Finally, for the question answering task, we propose that some of the earlier MUC resources could be reused. First, we assume that a MUC template defines the obligatory aspects of some topic description, and synthesize a narrative topic description from the template (a sort of ‘novelisation’ of the template). As in the preceding case an IE system could be benchmarked against a set of texts for which filled answer keys

exist by scoring its output answer keys using the MUC scoring software. It could then be run over the summaries produced by a summariser which has been given the full texts and the synthesized topic description. As with the previous cases, the difference between the benchmark computed on the full texts, and scores produced by automatically filling templates from the summaries should provide a measure of the information loss in the summary, and hence an indirect measure of its value.

## 5.2 An Initial Evaluation

We have not to date been able to carry out the proposed evaluations in full on our summarization system. In particular, since the goal of the basic summarization model we have implemented is to produce generic summaries, only the automated categorization evaluation is really appropriate.

However, since the MUC materials and scoring software were available to us, we were eager to attempt some form of an automated question answering evaluation, as described in the last section, using the existing LaSIE system. We therefore adopted a crude technique to simulate topic processing by the generic summarization system: the topic description (a TREC-style narrative) from the MUC-6 management succession IE task definition, was prepended, as the first paragraph, to each text. We then ran our summarisation system using the 'start of chain' criteria to select only those coreference chains which included a link between the topic description and the text. This scenario is similar to Baldwin and Morton (1998) who also use coreference between query, or topic description, and text to generate a user-focused, 'indicative' summary. Our approach, however, differs from theirs in that no special purpose mechanism is used to relate a query to a text, and also the detail of selecting sentences for inclusion in the summary is much simpler. Further, our evaluation is completely different in that they judge the summaries in terms of their capacity to support a relevancy judgment (i.e. the SUMMAC ad hoc task) whereas we evaluate the summaries in terms of their capacity to support a question answering-type task (MUC template filling).

Our technique was tried for 30 of the MUC-6 evaluation texts, and summaries produced using the length and spread criteria, with and without

the initial selection of focus chains. The full LaSIE MUC-6 IE system was then run with the summaries as input, and the resulting extracted templates scored. As a baseline, performance of the LaSIE system with the full texts as input was 45% recall, 64% precision, with 13 of 16 relevant texts correctly identified, by the production of a filled template, and 1 text proposed spuriously. Performance using the summaries from each set of criteria was:

Length of coreference chain:

6% recall 81% precision

4 texts relevant

17.7% of original text word count

Spread of coreference chain:

2% recall 89% precision

1 text relevant

15.6% of original text word count

Length of focus chain:

1% recall 80% precision

1 text relevant

6.1% of original text word count

Spread of focus chain:

1% recall 80% precision

1 text relevant

6.1% of original text word count

The relevance figures alone give some measure of the information loss between the full texts and the summaries, but the very low recall in all cases suggests that the summaries are really not suited to this task. One significant reason for this is that the MUC-6 topic description requires the identification of an event involving 3 instance types – an organisation, a management post and a person – while the production of the summaries is based on the selection of a single chain representing a single instance. The occurrence of all 3 required instances within the sentences of a single chain can therefore be expected to be rare. A more fruitful use of the topic description would be to produce a summary from all chains linking the text to the topic, rather than a single 'best' chain. Alternatively, for event-based topics, chains of event coreference relations could be used.

Out of interest, the evaluation was also run on summaries produced without the criteria requiring a link between the topic and the text, i.e. the generic summaries aiming to capture

the single main topic of each text. The summaries produced are considerably longer than with the topic description, but this gives considerably higher recall for the question answering task. However, given that the summariser here is attempting to capture the main topic of each text, independently of the question topic, these results are really a measure of the extent to which the text topics and question topic coincide in the corpus, rather than the suitability of the generic summaries for the question answering task.

Length of coreference chain:  
 33% recall 67% precision  
 10 texts relevant 2 spurious  
 87.0% of original text word count

Spread of coreference chain:  
 30% recall 65% precision  
 10 texts relevant 2 spurious  
 41.5% of original text word count

Length of focus chain:  
 19% recall 73% precision  
 8 texts relevant  
 19.6% of original text word count

Spread of focus chain:  
 19% recall 73% precision  
 8 texts relevant  
 18.2% of original text word count

A much more detailed analysis of these initial results is required before firm conclusions can be drawn, but the methodology does represent a useful step towards an automated evaluation procedure.

## 6 Conclusion

The use of coreference chains for the production of text summaries appears to be a promising technique. Further investigation into the most appropriate selection criteria for different purposes is required, and the current implementation is useful for such experimentation. To determine progress, however, automatic evaluation of summaries will be extremely valuable, and we have described initial experiments towards this goal using existing tools and resources. The current results demonstrate there may need to be subtle interactions between the selection of coreference chains and given topic descriptions, such as the use of event chains for

certain tasks, but the simple selection of a single chain may be still be useful as the basis for generic summaries.

## References

- S. Azzam, K. Humphreys, and R. Gaizauskas. 1998. Evaluating a focus-based approach to anaphora resolution. In *Proceedings of COLING-ACL'98*, pages 74–78.
- A. Bagga and B. Baldwin. 1998. Entity-based cross-document coreferencing using the vector space model. In *Proceedings of the COLING-ACL'98 Joint Conference (the 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics)*, Montreal, August.
- B. Baldwin and T.S. Morton. 1998. Dynamic coreference-based summarization. In *Proceedings of the Conference Methods in Natural Language Processing (EMNLP'98)*.
- R. Barzilay and M. Elhadad. 1997. Using lexical chains for text summarization. In *Proceedings of the ACL Workshop on Intelligent Scalable Text Summarization*, pages 10–17, Madrid, July.
- DARPA: Defense Advanced Research Projects Agency. 1998. *Proceedings of the Seventh Message Understanding Conference (MUC-7)*. Available at <http://www.muc.saic.com>.
- R. Gaizauskas and K. Humphreys. 1997. Quantitative Evaluation of Coreference Algorithms in an Information Extraction System. Technical report CS - 97 - 19, Department of Computer Science, University of Sheffield.
- R. Gaizauskas, T. Wakao, K. Humphreys, H. Cunningham, and Y. Wilks. 1995. Description of the LaSIE system as used for MUC-6. In *Proceedings of the Sixth Message Understanding Conference (MUC-6)*, pages 207–220. Morgan Kaufmann.
- M.A.K. Halliday and R. Hasan. 1976. *Cohesion in English*. Longman, London.
- I. Mani, D. House, G. Klein, L. Hirschman, L. Obrst, T. Firmin, M. Chrzanowski, and B. Sundheim. 1998. The TIPSTER SUMMAC text summarization evaluation: Final report. MITRE Technical Report MTR 98W0000138, MITRE.
- J. Morris and G. Hirst. 1994. Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics*, 17(1):21–45.
- K. Sparck Jones. 1993. What might be in summary? In Knorz, Krause, and Womser-Hacker, editors, *Information Retrieval 93: Von der Modellierung zur Anwendung*, pages 9–26. Universitätsverlag Konstanz.