Automatic Text Abstracting by Selecting Relevant Passages

MSc Dissertation in Cognitive Science and Natural Language

Submitted by
Klaus Zechner

Supervisors:

Steve Finch	Richard Shillcock
(HCRC, Edinburgh)	(CCS, Edinburgh)

August 12, 1995

To My Parents

Contents

0 Abstract **5**

1 Text Abstracting as a Task within Information Retrieval **5**

1.1 Some Key Concepts in IR 5
1.2 The Issue of Text Abstracting 7
1.3 Determining Relevant Passages in Documents: From Low Level to High Level Perspectives 8
1.3.1 Automatic Indexing: Determining Term Weights 8
1.3.2 Systems Using Term Statistics (and Heuristics) 9
1.3.2.1 The Abstract Generation System by Edmundson (1969) . . . 9
1.3.2.2 The Trainable Document Summarizer by Kupiec *et al.* (1995) 10
1.3.3 Systems Using Text Understanding 12
1.3.3.1 The FERRET IR system by Mauldin (1989) 13
1.3.3.2 The FASTUS information extraction system by Hobbs *et al.* (1992) 13
1.3.3.3 The BREVIDOC full-text retrieval system by Miike *et al.* (1994) 14

2 A System for Automatic Text Abstracting **15**

2.1 Design of a Simple Sentence Extraction System 15
2.2 System Description: Algorithm, Features and Usage 17
2.3 Results and Discussion 21

3 Experiment: Sentence Relevance Assessment by Humans **26**

3.1 Motivation and Design of the Experiment 26
3.2 Subjects 26
3.3 Results 27
3.4 Some Comments by the Subjects 27
3.5 Discussion 29

4 Comparison, Evaluation, and Discussion: Man vs. Machine **31**

5 Further Suggestions and Conclusions **36**

5.1 Possible Improvements of the System 36
5.1.1 Computational Issues 36
5.1.2 User Interface Issues 38
5.2 Suggestions for Additional Evaluations and Experiments 38
5.3 General Issues and Final Remarks 39

6 Summary **40**

7 References **40**

A The Sample Newspaper Texts **43**

B Results of the Automatic Text Abstracting System **50**

B.1 Abstracts for Article A (All 5 Methods) 50

B.2 Abstracts for All Articles (Method 3) 53

C Results of the Empirical Study 60

C.1 Significance Tables for Sentence Score Correlations 60

C.2 Significance Tables for Chi Square Evaluations 63

C.3 Optimal Abstracts According to Top Ranked Mean Scores 64

D Tables Concerning the Man–Machine Comparison 67

E README: Location, Description, and Usage of the Executables, Sources and Various Other Files 77

List of Figures

1 Precision–Recall Graph Comparing Methods 3 vs. 0 24

2 Precision–Recall Graph Comparing Methods 3 vs. 1 24

3 Precision–Recall Graph Comparing Methods 4 vs. 2 vs. 0 25

4 Instructions for Subjects in the Experiment. 27

5 Article A: Sentence Scores: Humans (Average) vs. Machine (Method 3) . . . 32

6 Article B: Sentence Scores: Humans (Average) vs. Machine (Method 3) . . . 32

7 Article C: Sentence Scores: Humans (Average) vs. Machine (Method 3) . . . 33

8 Article D: Sentence Scores: Humans (Average) vs. Machine (Method 3) . . . 33

9 Article E: Sentence Scores: Humans (Average) vs. Machine (Method 3) 34

10 Article F: Sentence Scores: Humans (Average) vs. Machine (Method 3) 34

List of Tables

1 Recall and precision of an example IR system, assuming 15 relevant items for a given user query. 7

2 Indicative data for the Daily Telegraph Corpus (approx. numbers are given). 21

3 Indicative data for the test articles. 22

4 Standard (default) parameter setting in the parameter file. 22

5 Article A: Selected sentences, ranks and weights for methods 0, 1, and 2. . . . 22

6 Article A: Selected sentences, ranks and weights for methods 3 and 4. 23

7 Precision, Recall and F scores for the five weighting methods (Article A). . . 23

8 Precision, Recall and F scores for the five weighting methods (all 6 test articles). 23

9 Significance of sentence score correlation: All 6 articles 28

10 Optimal selected sentences, ranks and scores for articles A, B, and C. 28

11 Optimal selected sentences, ranks and scores for articles D, E, and F. 28

12 Mean, minimum and maximum standard deviations for sentence scores. . . . 29

13 Mean Precision and Recall Across All Human Subjects (Articles A, B, C). . . 30

14 Mean Precision and Recall Across All Human Subjects (Articles D, E, F; All Articles). 31

15 Precision, Recall, Correlation and Closeness of Sentence Scores for Machine Method 3, Compared to Human Results. 35

16 Explanation of the "highest ranked word"–table. 53

17 Significance of sentence score correlation: Article A 60

18 Significance of sentence score correlation: Article B 60
19 Significance of sentence score correlation: Article C 61
20 Significance of sentence score correlation: Article D 61
21 Significance of sentence score correlation: Article E 62
22 Significance of sentence score correlation: Article F 62
23 Significance of sentence score chi squares (5/4 vs. 3/2/1): All 6 articles 63
24 Significance of sentence selections' chi squares: All 6 articles 63
25 Pearson's r and Precision (left below diagonal) for Article A. 67
26 Significance of Pearson's r and Recall (left below diagonal) for Article A. . . . 67
27 Pearson's r and Precision (left below diagonal) for Article B. 68
28 Significance of Pearson's r and Recall (left below diagonal) for Article B. . . . 68
29 Pearson's r and Precision (left below diagonal) for Article C. 69
30 Significance of Pearson's r and Recall (left below diagonal) for Article C. . . . 69
31 Pearson's r and Precision (left below diagonal) for Article D. 70
32 Significance of Pearson's r and Recall (left below diagonal) for Article D. . . . 70
33 Pearson's r and Precision (left below diagonal) for Article E. 71
34 Significance of Pearson's r and Recall (left below diagonal) for Article E. . . . 71
35 Pearson's r and Precision (left below diagonal) for Article F. 72
36 Significance of Pearson's r and Recall (left below diagonal) for Article F. . . . 72
37 Pearson's r and Precision (left below diagonal) for all Articles (first sentences are included). 73
38 Significance of Pearson's r and Recall (left below diagonal) for all Articles (first sentences are included). 73
39 Pearson's r and Precision (left below diagonal) for all Articles (first sentences are excluded). 74
40 Significance of Pearson's r and Recall (left below diagonal) for all Articles (first sentences are excluded). 74
41 Significance of Chi Squares: Sentence Scores: 5/4 vs. 3/2/1 (first sentences are included). 75
42 Significance of Chi Squares: Sentence Scores: 5/4 vs. 3/2/1 (first sentences are excluded). 75
43 Significance of Chi Squares: Selected Sentences (first sentences are included). 76
44 Significance of Chi Squares: Selected Sentences (first sentences are excluded). 76

0 Abstract

With increasing amounts of machine readable information being available one of the major problems for users is to sort out which texts are relevant for them and which are not. This paper is devoted to investigate methods of automatically creating text abstracts which ultimately can help users in deciding about which texts they are actually interested in and want to read in detail.

A system is designed and implemented which produces abstracts of newspaper articles from the Daily Telegraph Corpus (approx. 15 million words) by extracting a predefined (parameterisable) number of sentences of which a relevance-score is computed on the basis of combining the weights of the content words in the sentences.

For assessing the system's performance, an experiment is conducted where human subjects score the relevance of sentences of six articles and decide which of these they would select for an abstract.

It will be shown by means of statistical evaluation, that first, the system performs significantly better than a simple baseline case of selecting sentences from the beginning (and end) of the articles would do, secondly, that there is a highly significant correlation between the sentence scores of the system and the human subjects for the articles under consideration, and thirdly, that the system's precision and recall values are in the same range as precision and recall between the human subjects themselves.[1]

1 Text Abstracting as a Task within Information Retrieval

1.1 Some Key Concepts in IR

Textual information is stored in computer systems in at least two distinguishable forms:

1. in database management systems (DBMS): as data elements in tabular form, or
2. in information retrieval (IR) systems: as natural language texts in large collections or corpora.

Whereas for the former type, the user typically has to query the system in such a way that it can provide her[2] with the exact information she needs (*fixed ways of access*), things are different for the latter: With IR systems, what the user is looking for, is typically not a *specific* entry or item (or an array/table of these) but rather a list of items (in particular: documents, or parts of them, in certain circumstances) which "deal with" a subject matter she is interested in. Typical queries in IR systems consist of boolean AND–OR–NOT–combinations of keywords which the system then has to match to its database to retrieve a set of items as the "best approximation" to this query.

Hence, from a functional point of view (see Salton & McGill (1983: 10ff.)), an IR system consists of

[1] This abstract was *not* created automatically.

[2] Instead of the cumbersome he/she-notation, I will use the female referring pronouns throughout, for better readability.

- a set of information items (the data itself),
- a set of user queries or requests, and
- some mechanism which has to determine which of the information items match the user's request.

So there are — at least — three main tasks which an IR system has to accomplish:

1. indexing the information items, i.e. assigning a number of discriminative keywords (index words) to all items and (possibly) also weighting them[3]
2. mapping the user's request onto an internal representation language[4]
3. determining the proper subset of those information items the indexes of which have the highest similarity with the (converted) user query and return these items to the user

Usually, the user's initial query will be either too general or too specific or not "close to the point" (where the user wants to get but does not know how because she does not have the system's "map" of keywords). Therefore, the system's response will not satisfy her at first. Typically, in an interactive on-line system, an iterative process of successive refinement develops which will get shorter as the user gets accommodated to the system. Some more sophisticated IR systems are also able to learn from these adaptive processes and also might be able to help the user speeding up her search by giving her hints and suggestions about possible (related) keywords, combinations or request strategies in general.

In terms of assessment of the quality of IR systems, the following points are generally considered to be important (see e.g. Rijsbergen (1979: 145)):

1. the *user–interface*: how easy are access to and usage of the system, how fast is the learning curve?
2. the system's *speed*
3. the system's *recall*: the proportion of relevant items in the retrieved items
4. the system's *precision*: the proportion of retrieved items which are relevant

We shall explain the two terms *recall* and *precision* by giving a small example (see table 1). Assume, a collection holds n documents, $d_1 ... d_n$, which are indexed for a number of keywords contained therein. A user, being interested in information about "holidays in Spain", queries the system by using a boolean combination of index keywords, e.g., by formulating the query ("holiday" AND "Spain"). We now assume that 15 documents actually deal with "holidays

[3] This can be done and has often been done by humans but is increasingly done automatically, using statistical methods (see section 1.3.1).

[4] In general, this will have some form of the above mentioned boolean keyword-combination. However, the user does not always have to put her queries in this format directly; some IR systems allow for "higher level" input, in a format which is more like natural language.

Items Retrieved	Items Being Relevant	Recall	Precision
6	4	0.27 (4/15)	0.67 (4/6)
10	6	0.40 (6/15)	0.60 (6/10)
20	9	0.60 (9/15)	0.45 (9/20)

Table 1: Recall and precision of an example IR system, assuming 15 relevant items for a given user query.

in Spain", so these would be exactly the items the user is interested in. The system now performs a similarity computation and presents the 10 top ranked documents to the user. It turns out that 6 of them belong to the 15 "relevant" documents about "holidays in Spain" but 4 do not (e.g., they may contain sections about holidays and about Spain but none about holidays IN Spain, therefore getting a high similarity measure while not being relevant to the user.) The *recall* value is 0.4 (there were 6 out of 15 relevant items in the retrieved items), whereas the *precision* value is 0.6 (6 out of 10 retrieved items were relevant).

As the number of items grows which the IR system retrieves, recall improves but precision decreases: assume, we ask the system to present 20 items instead of 10. Say, there are now 9 relevant items in total (i.e., just 3 of the second 10 items are amongst the relevant ones), which yields 0.6 for recall and 0.45 for precision. The converse also holds, so if the system just presented 6 items, 4 of them being relevant, we get 0.27 for recall and 0.67 for precision. Both recall and precision are always in the interval [0.0,1.0], their optimum being at 1.0. But they are, as our example shows, inversely related to each other.[5] So, in a realistic IR system one has to find a compromise between these two parameters which can also be dependent on the user's interest: sometimes, when exhaustiveness is important, she will want to maximise recall, whereas if she exclusively wants to deal with *relevant* information, a high precision will be required.

A problem with the precision/recall-measure is that, in general, it is not a priori clear how actually to determine the subset of *relevant items* which is necessary for computing both of these indicators. One would have to ask the user to look over the entire document collection and to check each article in order to determine whether it would be relevant for her respective query or not. To avoid this laborious and realistically unmanageable task, there exist test–collections where the "optimal matching documents" for a given set of queries are known and the precision/recall values can therefore be computed.

1.2 The Issue of Text Abstracting

In many situations, users would and do prefer to look at text *abstracts* rather than at the whole text, before they decide whether they are going to read through the entire text or not. The two main types of abstracts are

1. (man-made) summaries
2. (machine-made) extracts

[5] Though this is an empirical observation and not a strict mathematical relation.

Man-made summaries, generally produced by the author herself, are very common in scientific articles and their purpose is to give the reader a *short* and *coherent* impression of the main idea of this article. Although, in general, the authors do no just collect the important phrases from their articles, Kupiec *et al.* (1995) showed that about 80% of the sentences in man-made abstracts were "close sentence matches", i.e., they were "either extracted verbatim from the original or with minor modifications." (Kupiec *et al.* (1995: 70))[6]

In a machine system, the goal of coherence is doubtless difficult to achieve and would (in the optimal case) require not only the system's ability of text understanding but also of text generation (see the next subsection). However, what a machine system can do quite readily, is to produce an *extract* of the text, i.e., to select a number of "most relevant sentences for this text" and present them to the user in text order. For many purposes, coherence is not a crucial point, but *relevance* certainly is. So: how does the system determine the relevance of passages in a given document? We will investigate this question in the following subsection, specifically by looking at a few existing systems in this domain, which range from "bottom-up" approaches, where a combination of statistical word distribution information and some either general or text specific heuristics are combined, to "top-down" approaches, where (at least *some*) understanding of the subject matter is crucial.

1.3 Determining Relevant Passages in Documents: From Low Level to High Level Perspectives

1.3.1 Automatic Indexing: Determining Term Weights

Since virtually all IR systems rely on some form of indices or keywords which are associated to all documents of a collection, it is a very natural step to think about computing the relevance of *parts* of these documents by employing this already stored index–information.

A standard way of generating text indices automatically is outlined, e.g., in Salton & Buckley (1990: 8):

1. Identify the individual text words
2. Use a dictionary of common high-frequency function words (i.e., a stop list) to ignore these words from the text
3. Use a suffix stripping routine to reduce the words to their stems
4. Compute a term weight for each word stem
5. Represent a document as a vector of the weights of all index terms in a collection (i.e., all word stems found by this procedure)

A standard method for term weight computation in IR is the "term-frequency * inverse-document-frequency" method (tf*idf, see e.g. Salton & Buckley (1990: 8)): The weight for a term k (w_k) is a product of the number of occurences of a term in a single document (tf_k)

[6] These abstracts were *not* created by the authors themselves but by professional abstractors.

times the logarithm of the number of documents in a collection (N), divided by the number of documents where this term occurs (n_k):

$$w_k = tf_k * \log \frac{N}{n_k} \tag{1}$$

This formula yields a high number for words which are frequent in one document but appear in very few documents only; hence, they can be considered as "indicative" for the respective document.

When a user's request is given to the system, it computes the similarity of the query-vector with all the document vectors and retrieves those documents with the highest similarity measure.

In the case of text abstracting, the IR system could use this information given by the user's query to produce an abstract of a document at the time of retrieval. This would mean to perform similarity measures for all the sentences, paragraphs or passages (depending on which of these vectors are available in the system) and return the highest ranked passages to the user.[7]

1.3.2 Systems Using Term Statistics (and Heuristics)

Interest in producing simple indicative abstracts, i.e., "extracts as abstracts", arose as early as the fifties, in the early days of (at this time fairly unsuccessful) machine translation. An important paper of these days is the one by Luhn (1959) who suggested to weight the sentences of a document as a function of high frequency words, disregarding the very high frequency common words (i.e., those in the above mentioned stop list).

1.3.2.1 The Abstract Generation System by Edmundson (1969)

Edmundson (1969) implemented a system for automatically generating text abstracts which, additionally to the standard keyword method (i.e., frequency depending weights), also used the following three methods for determining the sentence weights:

1. Cue Method: This is based on the hypothesis that the relevance of a sentence is affected by the presence or absence of certain (pragmatic) cue words; there were both "bonus" and "stigma" words in the cue dictionary, i.e., words which would have a positive or a negative effect on the respective sentence weight. These cue word lists were computed statistically, using man-made abstracts as a reference.

2. Title Method: Here, the sentence weight is computed as a sum of all the content words appearing in the title and (sub-) headings of a text.

[7] In our system, which is described in section 2, the underlying assumption is that there is no user query available, which means that the system has to produce the abstracting task just with the knowledge of the text itself and hence to rely on the computation of *all* the word weights of the document. However, one could also imagine an approach, where the title words or/and the words appearing in the first sentence (or: paragraph) could be used as a "simulated user query" but this would have to be done very carefully in order not to lose significant passages which do *not* contain words appearing in the title or first sentence/paragraph.

3. Location Method: This method is based on the assumption that sentences occuring in initial/final position of both text and individual paragraphs have a higher probability of being relevant.

The results showed, that the best correlation between the automatic and target (human-made) extracts was achieved using a combination of these three latter methods, i.e., omitting the key word method entirely. Since the key word method in isolation also has the lowest correlation coefficient of all the four methods, Edmundson (1969: 275) concludes that

> key words, although important for indexing, may not be as useful for extracting. This ... has important consequences for an abstracting system since avoiding frequency–counting the entire text results in considerable simplification and shorter running time for the computer program.

Several comments have to be made about this statement. First, it is very likely that an abstracting system nowadays would be embedded in an (standard) IR system, where the indexing already has been done and therefore is given for free. Secondly, all the other methods which were used by Edmundson (1969) are to a certain extent specific to the text sort and the text domain, i.e., they would have to be modified and adapted to a variety of different "text types" a large system has to handle. Thirdly, and maybe most importantly, it does not seem that Edmundson used an efficient algorithm for determining the key word weights, e.g., like the tf*idf one, described above. One cannot be sure about this though, because, unfortunately, Edmundson does not state the key word weight computation formula anywhere in his paper.

1.3.2.2 The Trainable Document Summarizer by Kupiec *et al.* (1995)

To mention a system presented very recently I shall shortly describe a trainable document summariser developed and described by Kupiec *et al.* (1995). Their system also performs a sentence extracting task, based on a number of different weighting heuristics. Specifically, the following features for sentence scoring, some of them resembling those employed by Edmundson (1969), were used and evaluated:

1. Sentence Length Cut-Off Feature: sentences containing less than a pre-specified number of words are not included in the abstract
2. Fixed-Phrase Feature: sentences containing certain cue words and phrases (or: following section headers containing these) are included
3. Paragraph Feature: this is basically equivalent to Edmundson's Location Method
4. Thematic Word Feature: the most frequent words (considering word length as parameter) are defined as thematic words; sentence scores are functions of the thematic words' frequencies
5. Uppercase Word Feature: upper-case words (with certain obvious exceptions) are treated as thematic words, as well

Kupiec *et al.* (1995) use a corpus which contains 188 document/summary pairs from 21 publications in the scientific/technical domain. The summaries were produced by professional abstractors and the sentences occuring in the summaries were aligned to the original document texts, indicating also the degree of "similarity": the vast majority (about 80%) of the summaries' sentences could be classified as "direct sentence matches", i.e., (almost) verbatim extractions or extractions with minor modifications only, having an equivalent content.

A statistical evaluation, which computes the performance for each of these features separately, shows that the summariser produces a recall between 0.2 and 0.33[8], the latter achieved by the Paragraph Feature. The best feature combination was (Paragraph + Fixed-Phrase + Sentence-Length) and yielded a recall of 0.44. Since the number of sentences produced by the system always equals the number of sentences in the corresponding manual summary, precision and recall are identical.

When looking at these results, one has, however, as Kupiec *et al.* (1995) mention, to take into account the inter- and intra-human variability, concerning the production of a document abstract. Rath *et al.* (1961)[9] showed in a study that extracts selected by four different human judges had only 25% overlap, and for a given judge over time only 55%.

Interestingly, the frequency-keyword features (the thematic word feature and the uppercase feature) also gave the poorest performance, similarly to the results of Edmundson (1969). One reason, as Kupiec *et al.* (1995) argue, might be found in the fact that these "general statistical" features tend to select sentences more evenly throughout a document, whereas in the corpus considered in their study, the important information is often located at the beginnings or ends. But the authors still do not drop these features because they argue that for a general-purpose-system one cannot rely for sure on this distribution of important material nor on the presence of certain indicative phrases. This argument was also important for the basic design decision of our newspaper article abstracting system which will be described in section 2.

Other methods and ideas which could be employed for automatically generated text abstracts in a bottom-up approach were suggested by various authors. I shall mention a few of them here:

- Text-Tiling: The principal idea behind text tiling, as described e.g. in Hearst & Plaunt (1993), is to identify coherent passages in a given text, to find "topic boundaries", which is done by computing similarity of adjacent blocks of texts.
- Hidden Markov Models (HMMs): As Mittendorf & Schäuble (1994) describe in their paper, HMMs prove to be a mathematically sound framework for document retrieval and also can yield effective retrieval results. If one approaches the task of text abstracting from a such a probabilistic modelling perspective, it might well be possible that HMMs could be employed for this purpose, as well.
- Clustering: Building links and/or clusters between index terms, phrases and/or other subparts of the documents has been employed by standard IR for a long time (see e.g.

[8] A relevant sentence is defined as either being a "direct match" or part of a "direct join", where a manually produced sentence was constructed by putting together a number of sentences from the original. In the latter case, *all* other members of this join must also have been produced by the system.

[9] Quoted from Kupiec *et al.* (1995).

Bookstein *et al.* (1995)); although this is not an issue in any of the above mentioned abstracting systems, it seems to be worth of consideration when building such systems (see also section 4).

1.3.3 Systems Using Text Understanding

The methods and systems we have looked at so far were purely statistical and heuristic in nature, they did not involve any understanding of the texts at all. However, as Mauldin (1989) points out, there appears to be sort of an "upper boundary" or limit for any system which solely uses keyword statistics (and, maybe, general heuristics). The main reasons of this "keyword barrier" are

- synonymy
- polysemy
- anaphora
- metaphors, metonymy
- context sensitivity

This means, that various forms of ambiguity, together with the problem of anaphora resolution must be dealt with, if one aims to achieve a system which performs significantly better than a purely statistically based keyword and heuristics approach would do. But here we are, in the heart of the "big issues" in computational linguistics and natural language understanding; some of these problems tend to be "AI complete", i.e., they would require the existence of a "global and common sense world knowledge" component to sort out all possibly occuring ambiguities in the analysis process. The complexities and intricacies of this task are the major reasons why all high-level systems built so far had to focus on a fairly restricted domain which at once exclude them for dealing with general domain texts, as it is necessary, e.g., for newspaper articles.

Another important point to mention is that these high-level systems are, in general, not designed for producing text abstracts.[10] Most of them are information extraction systems, where entries or templates in some sort of database have to be filled while scanning and processing the respective texts. Although several authors state that "in principle" their systems could be used for the purpose of text summarisation, there is the not too insignificant gap of text generation which would have to be considered carefully to produce a sensible and coherent output.

In this section, we will briefly look at the following three systems:

- the Ferret system by Mauldin (1989)
- the FASTUS system by Hobbs *et al.* (1992) and
- the BREVIDOC system by Miike *et al.* (1994)

[10] An exception is the BREVIDOC system described below.

1.3.3.1 The FERRET IR system by Mauldin (1989)

The FERRET information retrieval system, as it is described in Mauldin (1989), consists of the following three components:

- a **text parser** which reads newsgroup texts (from an astronomical newsgroup) and converts them into symbolic representations that are stored in a database
- a **query parser**[11] which serves as a user interface to generate case frames patterns for the
- **case frame matcher** which brings the two parses together and determines whether a given case frame pattern matches a representation generated by the text parser

The case frames which are generated by the text parser are an uninstantiated script[12] and a "conceptual dependency graph"[13].

An important feature of the parser[14] is that it does not aim at a "full" parse but uses a "text skimming" method which can extract the general meaning of the text, ignoring or suppressing some of its details. It deals with partially built structures and tries to find all the possible meanings of a sentence so that the FERRET system can assign ambiguous text to two or more categories; the ambiguities are resolved by using expectations based on the scripts generated by the parser.

In a comparison study, Mauldin showed that the retrieval performance of the FERRET system, in terms of recall and precision, was significantly better than in a standard IR approach which uses a statistically based boolean keyword search.

1.3.3.2 The FASTUS information extraction system by Hobbs *et al.* (1992)

The FASTUS system, developed by Hobbs *et al.* (1992) at SRI, is built for extracting information from "free text" for entry into a database. The domain it operates in, are newspaper articles about Latin American terrorism.[15] In contrast to *text understanding* where the aim is to get a target representation which accommodates the full complexities of language, *information extraction* has somewhat more moderate goals (Hobbs *et al.* (1992: 4)):

- only a fraction of the text is relevant (in the case of the MUC-4 texts, only about 10% is relevant)
- information is mapped into a predefined, simple, and rigid target representation

[11] Actually, this parser was not implemented at the time of Mauldin's paper but was simulated instead by the text parser.

[12] For the notion of *scripts* see e.g. Lehnert (1980).

[13] This is derived from the "Conceptual Dependency Theory", developed by Schank *et al.* (1975). An introduction into the concept of *frames* can e.g. be found in Minsky (1980).

[14] A re-implementation of the FRUMP parser, see DeJong (1979).

[15] This is due to the MUC-4 evaluation of text-understanding systems (see Sundheim (1992)) where SRI participated in.

- subtle meaning nuances and writer's intentions can be disregarded entirely

FASTUS is designed essentially as a set of cascaded, nondeterministic finite state automata which makes the system considerably fast.

The FASTUS system's operation consists of the following four steps (there are preprocessing and post-processing procedures, as well, to do initial format standardisation, spelling correction and the final conversion of FASTUS's incident structures into the format required by MUC-4 templates):

1. Triggering: the system looks for certain trigger words and matches the phrase to a predefined pattern
2. Recognising Phrases: this comprises the detection and construction of noun groups, verb groups, and the detection of prepositions and particles
3. Recognising Patterns: from the output of the previous step, the patterns of interest are identified and corresponding "incident structures" are built up which represent certain real-world states or events described in the article
4. Merging Incidents: those incident structures that derive from the same incident are identified and merged together in order to form the actual database entries

Although this system basically faces the same major drawback as all the high-level systems, namely that it handles a very limited domain of texts only, the authors argue that due to the speed of the system they are (and actually were) able to adapt FASTUS to a new domain knowledge in a very short period of time.

1.3.3.3 The BREVIDOC full-text retrieval system by Miike *et al.* (1994)

Miike *et al.* (1994) present a system named BREVIDOC which is a full-text retrieval system that enables the user to specify an area within the text for abstraction and also to control the volume of the abstract interactively. BREVIDOC builds heavily on Rhetorical Structure Theory (RST), developed by Mann *et al.* (1989). The heart of the system is the Document Structure Analyser which performs the following processes:

1. Document Organisation Analysis: This process determines the "outer" structure of a document, i.e., the division into sections, their headings, and the paragraph structure of the sections.
2. Sentence Analysis: Here, the sentences are parsed by a morphological and a syntactic analyser.
3. Text Structure Analysis: This process first extracts the rhetorical relations between sentences, by matching them against a set of connective expressions and building a sequence of sentence identifiers and relations. Then, a number of *if-then*-rules, which look at certain cue words and patterns, are applied for segmenting these sequences. Finally, binary trees which represent the rhetorical structure of the text are constructed and

heuristic rules are used to select the most plausible tree candidate. This whole analysis is done for separate paragraphs first and then is performed again for inter-paragraph structures, using the rhetorical relation of the first sentence of each paragraph.

4. Semantic Role Extraction and Word Indexing: For each sentence in every document, a semantic role is extracted, by matching sentence-trees to *if-then*-rule patterns. Eventually, for every semantic role, a word index is generated that is based on document IDs and words obtained by morphological analysis of sentences with that role.

When an abstract is to be produced, the abstract generation function looks at the analysed text structure for each section, and then generates an abstract based on the relative importance of the rhetorical relations which were found therein. The system uses cumulative penalty scores for different types of relations, depending on whether the respective node is belonging to a right-, left-, or both-nucleus-relation and on which branch of the subtree it occurs.

Although the results of preliminary experiments suggest that this approach is quite useful both in identifying the correct structures of the texts and in selecting the most important key sentences for the abstracts, there is again some restriction in terms of the possible text domains, since the rules for semantic roles were constructed for the purpose of analysing technical papers. So, when the authors state (Miike *et al.* (1994: 152)) that their system uses "linguistic knowledge only and is thus domain-independent," the proof for this claim has yet to be done. Another argument which would weaken this statement is that there are text sorts where rhetorical structures are much less explicit than in technical articles (or scientific journal articles), e.g., texts whose main purpose is a narrative one. In this case, the system barely can rely on the sentence connectives and certain cue words or patterns but certainly has to use other, more low-level methods, as well.

2 A System for Automatic Text Abstracting

2.1 Design of a Simple Sentence Extraction System

Following the lines of traditional IR approaches, a text abstracting system is designed which should be able to generate "extracts as abstracts" by selecting relevant sentences from newspaper texts ranging over arbitrary domains. This "general domain" restriction rules out most of the high-level designs that were discussed in the previous section. Instead, the system relies primarily on statistically derived word-weight assignments and computations, with minor tunings and adjustments due to some of the intrinsic properties of newspaper articles. The principal idea behind the implementation of the system is to determine how much a simple and effective standard IR method can achieve for the task of automatically constructing text abstracts in unrestricted text domains.

The main steps which the system has to perform are the following:

1. take an article from the corpus which is marked up in SGML format (see Goldfarb (1990)) for sentences and words

2. build a word-weight matrix for all content words in the article, excluding high-frequency words given in a stop-list and using the tf*idf weight measure where the document frequencies of all the words are given in a precomputed file

3. determine the sentence weights for all sentences in the article by combining the word weights and using additional features and/or heuristics, specific to newspaper articles
4. sort the sentences according to their weight and rank them
5. return the highest weighted sentences below a (pre-specified) threshold-rank (which possibly may be slightly shifted in either direction) to the user in text order

Before actually starting with the implementation of the system, about thirty articles with a minimum length of about 500 words from the Daily Telegraph Corpus[16] were manually abstracted in order to get some initial ideas what a computerised abstracting system should or could "take into account" when performing its task on this newspaper corpus. There were a few observations made which I shall mention briefly:

- the words appearing in the title often are good indicators for the relevant words in the article as a whole; though these words cannot, of course, be used exclusively for the purpose of relevance determination
- the information given in the first few sentences (sometimes: the first paragraph) of an article are highly indicative for its content most of the time; relevant passages also tend to occur in the last few sentences but this is only a slight trend
- for articles which contain very "dense" information, especially for some of those in the domain of business and finances, it is often difficult to produce an abstract *at all*, without omitting fairly relevant information; this restricts the previously claimed domain-independence to a certain extent but it can be argued that in these cases the abstracting task would be hard to perform for humans as well
- producing an abstract turns out to be more difficult in multi-topical articles where topics are "merged" as opposed to mono-topical ones (which are the standard case, at least for most of the newspaper articles in this corpus); this problem of dealing with texts which are about more than one (main) topic is of course a general and an important one; however, it was not addressed in the present system since it does not seem to be very central for the newspaper articles in this corpus

These observations gave rise to the introduction of the following heuristic tuning parameters:

1. Title Word Factor: For every (content) word in the article which also appeared in the title, the weight is multiplied by a user-defined factor
2. Sentence Factors: For sentences occuring in the initial or final section of an article, their weight can be increased (or decreased) by user-defined factors

Instead of using a word-stemming algorithm (which would certainly be preferable but was not available to the author), a simple word-length-cut-off procedure was used: It was assumed that, in general, the first six letters of a word are "indicative" of its content. This method

[16] This corpus contains approximately 15 million words. It served as a database for the system throughout.

has the advantage of reducing the global word list considerably and thereby speeding up the system, but of course the disadvantage that word level ambiguities increase at the same time. In the experiments, however, this did not turn out to be a major problem but first, we looked at very few articles only, and secondly, the articles in the corpus are fairly short and hence barely give rise to ambiguity problems of this kind.

Concerning other suggestions for system features and heuristics from the literature which were discussed above, it turned out that a few of them were readily applicable to the newspaper articles of our corpus; this is true for the features *location*, *title* and *key word*. However, concerning the *cue word method*, it is doubtful whether many cue words can be sorted out which can perform equally well on newspaper articles as they would do for other, commonly considered domains, like scientific journals. However, if one had a fairly large training set of human-made abstracts for this corpus, one could try to clarify these matters statistically, as Edmundson (1969) did in his work.[17]

Another feature, employed by Kupiec *et al.* (1995), namely the *sentence length cut-off feature*, was also implemented, in two different variants. Their *uppercase word feature* was, however, not implemented for the simple reason that the whole system works with upper-case words only and therefore cannot get hold of this distinction at all.

Paragraph information is disregarded for this system for two main reasons: first, the typical length of a paragraph is just about two sentences in this corpus and secondly, the tool available for marking up sentences in the corpus eliminates the paragraph information.[18]

2.2 System Description: Algorithm, Features and Usage

The text abstracting system was programmed in C++, using specific libraries for both the handling of files in SGML-format and for the management of hash tables used for fast access to the (content) words.

The algorithm performs the following steps the more important of which will be described in detail below:

1. initialise hash tables: stop word table, word table
2. read in the parameter-file, containing user-defined values for various system features (e.g., weight factors, output format parameters)
3. read in the stop word list from a file and enter them to the stop word table
4. read in an article from the corpus
5. for all sentences in the article do:
    (a) get the next word and convert it to upper case (punctuations are ignored)

[17] Maybe it would turn out that there were more "stigma" than "bonus" cue words, stigma words being e.g. personal comments by the authors, stating their point of view.

[18] This paragraph information would have had to be either included manually or a separate tool for marking up sentences *and* paragraphs would have had to be constructed.

(b) check: if the word is in the stop word table, drop it and goto 5a

(c) ... else: truncate the word to its first six characters (this word will be synonymously referred to as "content word", subsequently)

(d) check: if the word is already in the word table, increment both the word-counter for the current sentence and the global word counter and goto 6

(e) ... else: add this word to the word table and initialise the word-counters

6. read the corpus word list file to get the word-in-document frequencies for all the words in the word table

7. compute the tf*idf weights for every word

8. for all five sentence weighting methods do:

   (a) compute the sentence weights for every sentence

   (b) sort the sentences according to their weights

   (c) shift the threshold-rank if appropriate

   (d) normalise the weights for comparison with the results of the experiment (see section 3)

   (e) read the article file again and output the title and every sentence whose rank is below or equal the threshold-rank

To Step 2: The user can set a few parameters herself in the parameter file which the system reads at the beginning of the abstracting procedure:

- Maximum Word Length: For all our experiments, this was set to 6, but it can be changed to any desired length.[19]
- Title Factor: This factor is applied to content words appearing in the title; experiments showed that a good value for this factor is around 1.5.
- Sentence Weight Factors: The weights of document-initial and -final sentences (maximally 10 each) can be tuned by these factors.[20] In testing various factors it turned out that it useful to use moderate factors for the first two or three sentences (around 1.4) and possibly *very* moderate ones (around 1.1) for the last few sentences of the article.[21]
- Default Threshold Rank: This determines the number of sentences which are extracted. (For the comparison between humans and machine results, this parameter was set to 6 since humans were to select 5–7 sentences from each article.)

[19] One has to bear in mind, however, that one would have to generate a new global word list (with the word-in-document frequencies) if one wanted to change this parameter.

[20] The title is disregarded here since it will be included in the abstract automatically.

[21] To ensure that the last sentence which contains the non-informative copyright notice is not included in the abstract, its factor was set to 0.05.

- Threshold Move Factor: The default threshold rank will be shifted (maximally by 1 up or down) if the ratio between neighbouring weight differences exceeds this number. The purpose of this (minor) adjustment is to *exclude* sentences occuring after a "weight-gap" in the sorted sentence weight list, and to *include* sentences occuring before such a gap. If both conditions are met, priority is given to the latter (i.e., inclusion of another sentence which corresponds to increasing the threshold-rank by 1) since this yields a better recall which is in general preferred in this abstracting generation task.[22]

- Sentence Length Cut-Off: For method 4: This specifies the minimum number of content words in a sentence to be included in an abstract. For method 2: Here, it specifies the minimum divisor in the sentence weight formula (see below) which affects only sentences which have fewer content words than this number.

- Minimum Sentence Number: If an article has fewer sentences than the number specified here, no abstract is produced — it is assumed that the effort of abstract generation is not worthwhile comparing to just presenting the entire article to the user.

In addition to these parameters which affect the actual task of abstracting, there are a few switches and parameters for customising the text output of the system, e.g., whether and how many of the highest ranked words shall be shown, whether and how the texts of the abstracts shall be printed, whether and in which format information about sentence-weights and -rankings shall be given in the output file.

To Step 3: The stop word file was created by taking about 50 words from the corpus word file which appeared most frequently; these are mainly closed class words like determiners, conjunctions, auxiliaries, pronouns, and prepositions. In total, this word list covers approx. 40% of the total word occurences in the corpus.[23]

To Step 6: In this file, the (truncated upper-case) words occuring only once were omitted, reducing the file-size by about 30% and speeding up the system.[24]

To Step 7: The formula for determining the word weight w_i is the standard tf*idf one, except that a scaling factor in the argument of the logarithm was introduced to get resulting numbers in a "sensible" range (i.e., numbers which are not too close to zero for words which have a high word-in-document frequency):

$$w_i = tfact_i * \mathtt{f}doc_i * \log \frac{100 * D}{\mathtt{f}art_i} \tag{2}$$

where $tfact_i$ is a user-defined title-factor for increasing the word-weights for words (also) appearing in the title, $\mathtt{f}doc_i$ the number of occurences of the word in the article (document),

[22] There are no threshold adjustments with method 1 since the sentence weights are irrelevant there.

[23] The stop words are: THE OF TO LT A AND IN IS FOR THAT ON BE WAS BY WITH AT IT ARE AS HAVE FROM HAS WILL NOT WHICH AN F ITS BEEN HAD WERE WOULD THEIR BUT THEY PLC C WHO THIS OR UP WHAT DO WHERE THERE ANY HE HIS MR SAID TOLD SAY DID

[24] A default value of 1 has to be set for all content words in the end anyway, for "security" reasons, i.e., in order not to run the risk of a "division by zero" error in the idf-computation.

A further speed improvement would be to store the idf-values directly in this corpus word list file instead of computing them in the program itself.

D the number of documents in the corpus and $\mathtt{f}art_i$ the word-in-document frequency: the number of different articles where the word occurs.[25]

To Step 8a: For computing the sentence weights s_w, the following five different methods are implemented; in each of these methods, the scaling factor $sfact$ is a user-defined sentence-weight factor, N is the number of content words in the sentence and A is the scaled number of documents in the corpus ($100*D$):

- Method 0: summing over all content words in the sentence and building the average by dividing the sum by the number of content words:

$$s_w = \frac{sfact}{N} * \sum_{i=1}^{N} tfact_i * \mathtt{f}doc_i * \log \frac{A}{\mathtt{f}art_i} \tag{3}$$

- Method 1 ("default"): just choose sentences from the beginning and end of the article, ignoring all the word weights: choose at least 3 sentences from the beginning and at most 3 sentences from the end; the algorithm for this method has to determine the "front-limit" ($flimit$) and the "back-limit" ($blimit$), i.e., the sentence indices for the beginning-part and the end-part, respectively (the total number has to be equal to the user-defined number of sentences which are to be extracted ($numsent$)):
  - if $numsent <= 3$: $flimit = numsent$; $blimit = 0$
  - if $4 <= numsent <= 6$: $flimit = 3$; $blimit = numsent - 3$
  - if $numsent > 6$: $flimit = numsent - 3$; $blimit = 3$

- Method 2: short sentences get less weight than with method 0; this is achieved by dividing the word-weight sum by a user-defined number (lim) *if* this is larger than the actual number of content words in the sentence:[26]

$$s_w = \frac{sfact}{\max(lim, N)} * \sum_{i=1}^{N} tfact_i * \mathtt{f}doc_i * \log \frac{A}{\mathtt{f}art_i} \tag{4}$$

- Method 3: summing over all content words in the sentence; no average is built here:

$$s_w = sfact * \sum_{i=1}^{N} tfact_i * \mathtt{f}doc_i * \log \frac{A}{\mathtt{f}art_i} \tag{5}$$

- Method 4 ("sentence length cut-off"): all sentences having less content words than a user-defined number lim are set to 0, else the formula of method 0 is used

To Step 8d: Since the subjects in the experiments were to use (discrete) sentence scores from 1 to 5, the resulting scores of the machine system were also mapped to these numbers in the following way: First, the sentence indices for the maximum and minimum weights are

[25] Since all these computations are based on six-character upper-case words, the respective document frequencies were, of course, also computed for these stripped words.

[26] Hence method 2 lies "in between" methods 0 and 4.

total words	15.000.000
different words	180.000
different truncated upper-case words	80.000
articles	44.000
average words per article	340
average sentences per article	17
average words per sentence	20

Table 2: Indicative data for the Daily Telegraph Corpus (approx. numbers are given).

determined.[27] Then, the actual sentence weights are linearly transformed into the interval [0,5] and each sub-interval [0,1], [1,2] etc. is mapped onto its upper boundary.

The system is invoked via the command `doprecis*` with the name of the input file as a first, and the name of the output file as a second parameter, respectively.[28] The input file has to be a single article in SGML format which is marked up at least for sentences and words.

The system was developed and tested on a SUN sparc workstation under Unix (Sun OS 4.1.3); the typical run-time for processing an article of the test set (average length about 550 words) was about 3 seconds (user-time). If multiple articles were to be processed in a sequence, the average speed could be increased significantly by reading in the corpus word list file *first* and processing the documents one after the other, having initialised the word-counters every time before a new document was to be processed.

2.3 Results and Discussion

For testing the system and for comparing the system's output to the abstracts produced by humans in the experiment described in section 3, a set of 6 articles was chosen from the corpus which fulfilled the criteria of (a) having more than average article length and (b) not belonging to the domain of business/finance for reasons explained earlier. These texts are given in full in appendix A.

Some indicative data about the Daily Telegraph Corpus is given in table 2, the data for the six test articles in table 3.

We will now briefly look at some sample results of the system when working with Article A as an input file, using the "standard" parameters in the parameter-file (see table 4).

The relevant parts of the output file, including ranking and weighting information and abstracts for all of the five different methods are presented in appendix B. Tables 5 and 6 show

[27] If the title sentence has maximum weight, it is disregarded, since the title was not scored in the experiment and secondly, its weight can be considerably high compared to the rest of the sentences which is due to the Title Weight Factor.

Similarly, if the last sentence, containing the copyright-information, is least ranked, it is disregarded, as well, since it will usually get a very low sentence weight.

[28] If the second parameter is omitted, the system produces an output file named as the input file with a suffix `.out`.

Article	Words (including title)	Sentences (including title)
A	508	20
B	608	22
C	570	24
D	534	22
E	556	22
F	551	21

Table 3: Indicative data for the test articles.

Title Factor	1.5
Sentence Factors for Beginning	1.4 1.4 1.2 1.1 1.0 ... 1.0
Sentence Factors for End	0.05 1.1 1.1 1.1 1.0 ... 1.0
Maximum Word Length	6
Default Threshold Rank	6
Threshold Move Factor	2.5
Sentence Length Cut-Off	8
Minimum Sentence Number	10

Table 4: Standard (default) parameter setting in the parameter file.

the sentence numbers[29], their ranks and scores[30] for all five methods.[31]

In order to be able to compare the results of these five different abstracting methods, we will look at their respective precision-recall graphs for various default threshold ranks (i.e. numbers of extracted sentences) for all six test articles together (figures 1 to 3). The "relevant" sentences are taken to be those which the human subjects chose in the experiment described in the next section. All precision-recall values are means over all 13 human subjects.

In table 7, precision/recall values are given for the five weighting methods, considering article A and the case with standard parameters; an additional measure for precision-recall, the

[29]The title is excluded in these tables since it is present in all of the abstracts anyway.

[30]Except for method 1, since for this method, the rankings and weights are irrelevant.

[31]Note that the sentence numbers differ from those in the output file; in fact they correspond to the text-numbering in the experiment (which is also the numbering used in the full text listings in appendix A). There, the initial author's line was disregarded as well as the subtitle in article E; hence, the numbering differences are 1 for articles A–D and F, resp. 2 for article E.

Method 0			Method 1	Method 2		
number	rank	weight	number	number	rank	weight
1	1	39.6	1	1	0	39.6
3	2	28.2	2	3	2	28.2
4	4	25.9	3	4	4	25.9
6	3	27.9	15	6	3	27.9
9	5	23.6	16	9	5	23.6
13	6	23.4	17	13	6	23.4

Table 5: Article A: Selected sentences, ranks and weights for methods 0, 1, and 2.

Method 3			Method 4		
number	rank	weight	number	rank	weight
1	0	792	1	0	39.6
2	3	532	3	1	28.2
3	5	452	4	3	25.9
4	2	543	6	2	27.9
6	1	559	8	6	22
7	4	482	9	4	23.6
12	6	412	13	5	23.4
			14	7	22

Table 6: Article A: Selected sentences, ranks and weights for methods 3 and 4.

Method	Precision	Recall	F ($\beta = 0.5$)	F ($\beta = 1.0$)	F ($\beta = 2.0$)
0	0.31	0.33	0.31	0.32	0.33
1	0.49	0.5	0.49	0.49	0.5
2	0.31	0.33	0.31	0.32	0.33
3	0.48	0.6	0.5	0.53	0.57
4	0.34	0.47	0.36	0.39	0.44

Table 7: Precision, Recall and F scores for the five weighting methods (Article A).

F-score, is computed according to the formula (see Hobbs *et al.* (1992: 4)):

$$F = \frac{(\beta^2 + 1)PR}{\beta^2 P + R} \tag{6}$$

where P is precision, R is recall and β is a parameter encoding the relative importance of precision and recall: If $\beta = 1.0$, they are weighted equally, if $\beta > 1.0$, recall is more significant, if $\beta < 1.0$, precision is.[32] In table 7, values are given for $\beta = 0.5, 1.0$ and 2.0.

Table 8 is analogous to the previous table except that the computation was performed for all the six test articles together.

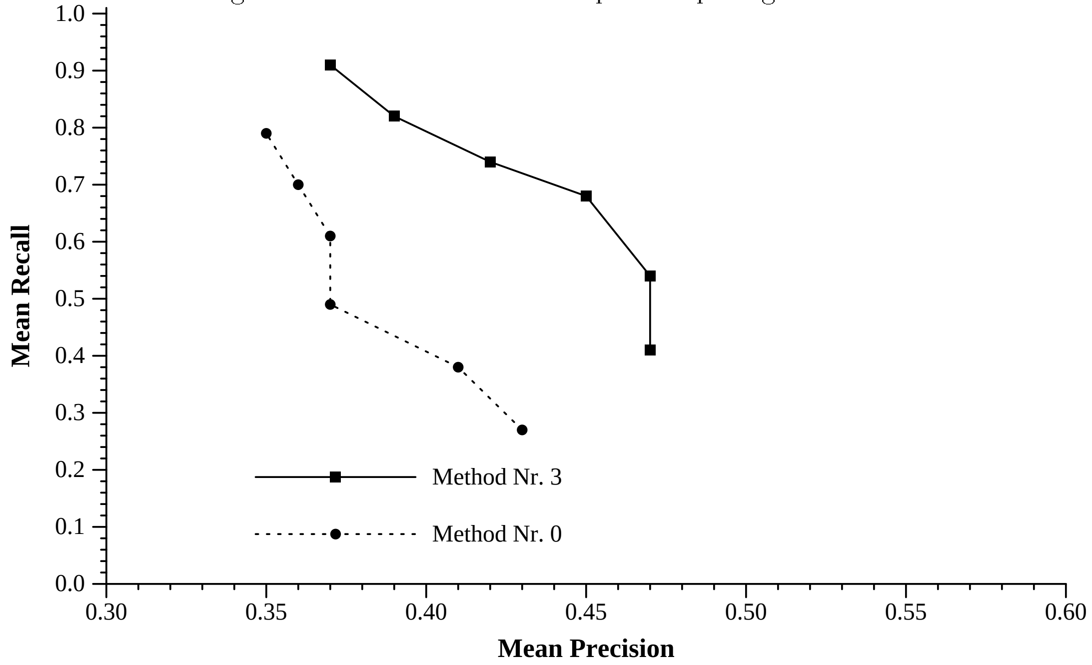
The numbers and graphs for precision and recall clearly favour method 3 which is just the simple summation method. Also, as we will see later, this method yields the best correlation with the human subjects in terms of sentence scores.

[32] This is mixed up in the paper by Hobbs *et al.* (1992) erroneously.

Method	Precision	Recall	F ($\beta = 0.5$)	F ($\beta = 1.0$)	F ($\beta = 2.0$)
0	0.41	0.38	0.4	0.39	0.39
1	0.44	0.4	0.43	0.42	0.41
2	0.41	0.41	0.41	0.41	0.41
3	0.47	0.54	0.48	0.5	0.52
4	0.4	0.48	0.41	0.44	0.46

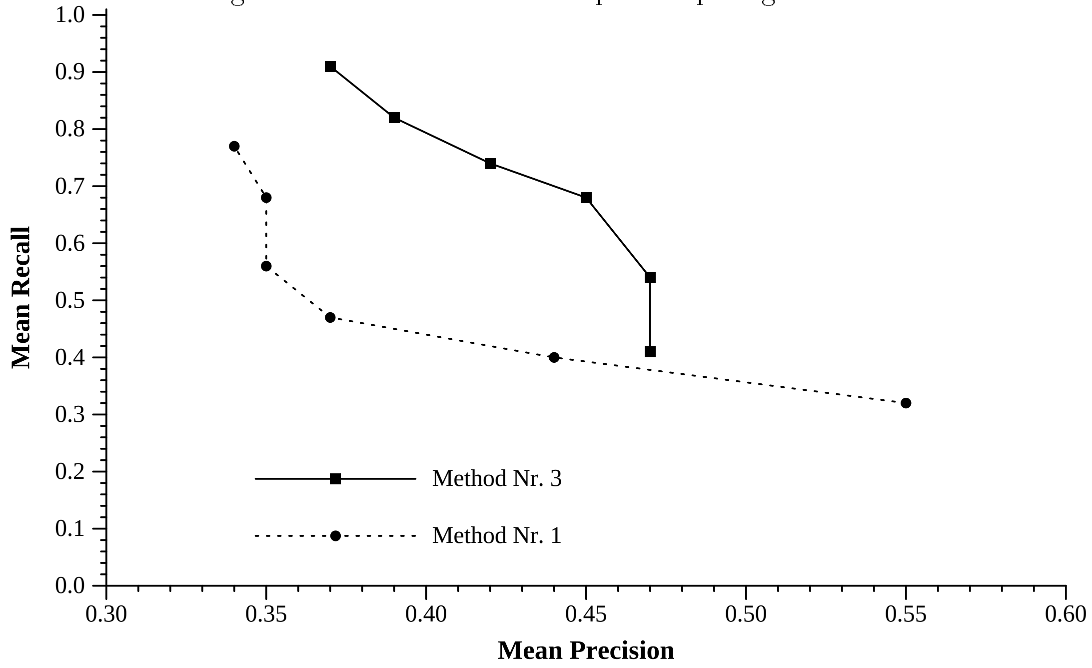
Table 8: Precision, Recall and F scores for the five weighting methods (all 6 test articles).

Figure 1: Precision–Recall Graph Comparing Methods 3 vs. 0



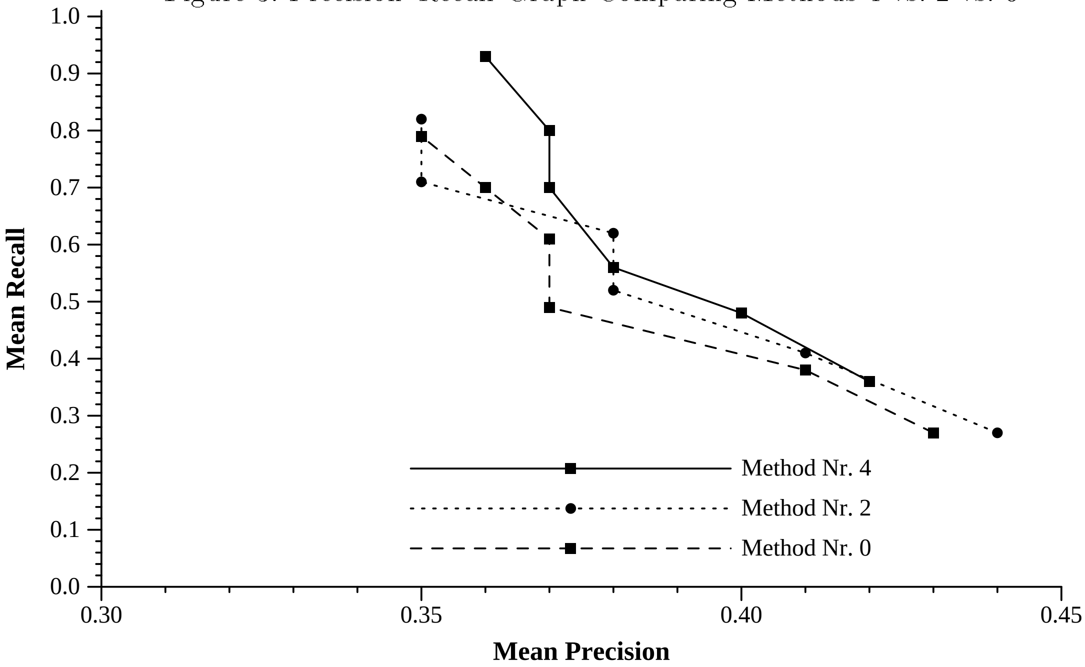
Parameter: Number of Sentences: 14 - 12 - 10 - 8 - 6 - 4

Figure 2: Precision–Recall Graph Comparing Methods 3 vs. 1



Parameter: Number of Sentences: 14 - 12 - 10 - 8 - 6 - 4

Figure 3: Precision–Recall Graph Comparing Methods 4 vs. 2 vs. 0



Parameter: Number of Sentences: 14 - 12 - 10 - 8 - 6 - 4

Unfortunately, the test corpus, consisting of just six articles, is far too small for an automatic trimming and tuning of the various system's parameters. However, the effort of obtaining large samples of human-made abstracts was beyond the scope of this work. As for the title factor and the sentence factors, a number of different values were tested manually and the (in this case intuitively) best values were set as the above listed default parameters in the parameter file. There were three tests performed on the sentence cut-off factor: it was assigned to the values 5, 8 and 11 respectively.[33] But it turned out that the precision/recall and correlation values differed only marginally for different sentence cut-off factors. In fact, both methods using this factor (method 2 and method 4) are less successful than method 3.

As it can be expected, method 1, the "default" method, produces significantly weaker results — it does not involve any sentence weighting and ranking at all.

An explanation why building the average in method 0 is somewhat worse than just taking the sum of the word-weights as the sentence weight (method 3) might be that in general, a lot of medium and low weighted words would "cancel out" the effect of high-weighted words if the sum was divided by the total number of content words in that sentence. This would specifically be true for longer sentences where the ratio of "relevant vs. irrelevant" words becomes smaller. Another explanation would be that humans tend to pick out long sentences rather than short ones and a method which normalised the sentence weight would hence be significantly disfavoured.

In section 4, we will look at some details of the system's performance again, specifically at the comparison between the top method 3 and the results of the experiment whose description follows in the subsequent section.

[33] Kupiec *et al.* (1995) used the value 5 for their system.

3 Experiment: Sentence Relevance Assessment by Humans

3.1 Motivation and Design of the Experiment

There were two main motivations for carrying out an experiment which would involve human subjects doing text abstracting: First, to investigate whether and in how far there is a "coherence" in this task, i.e., whether there exists, at least to some extent, an "ideal abstract" which different subjects agree about. And secondly, to get a database for the assessment of the computerised abstracting system: how closely the machine's results match those of the subjects, how significant is the correlation between human- and machine-produced abstracts?

For the experiment, six articles from the Daily Telegraph Corpus were chosen, according to the above mentioned criteria (see appendix A). Three of them (articles A, D, F) were presented in their normal format, as they would appear in a newspaper, i.e., with bold titles and paragraphs being indented. The others (articles B, C, E) were presented without the title (and, for article E, without its sub-title) and without any paragraph information. The reason for this was to find out whether there were greater variabilities in the latter group, assuming that title and paragraph information would help the subjects a lot in mentally structuring the article and therefore lead to more conformity in the abstracts in the former group.

Subjects should perform these two tasks for each of the six articles:[34]

1. give every sentence a relevance score between 1 (barely relevant) and 5 (highly relevant), and

2. select a number of about 5 to 7 sentences to be included in the abstract

The contents of the instruction sheet is given in figure 4.

3.2 Subjects

The experiment sheets were handed out to 15 subjects, all native speakers of English and most of them postgraduate students; subjects were paid for the experiment. 14 set of sheets were returned. On average, it took the subjects about 50 minutes to complete the task. Two subjects forgot to carry out instruction 3 (i.e., circling the sentences which they would select for the abstract) in which cases the "default" sentences, i.e., those ranking highest were assumed to be the selected ones.

The global analysis showed a surprisingly good correlation across all subjects (see table 9). However, subject nr. 7 did not correlate with 8 of the remaining 13 other subjects. There were indications that his results were generally rather produced in a "rough-and-ready" manner which gave rise to the decision to exclude him from the subsequent analysis; so there are in total 13 subjects whose results were compiled and evaluated subsequently.

[34] The articles were randomised in their order for every subject.

Your task is to read through the following six newspaper articles, and for each article to choose a number of about **5–7 sentences which are most central (or relevant) for the meaning of the article**. That is, that these sentences together would form sort of a "primitive abstract" of the respective article.
Please proceed, for each article, in the following way:

1. Read through the article to get an idea about its content.
2. Read it a second time, sentence by sentence, give each sentence a **score of relevance between 1 (barely relevant) and 5 (highly relevant)** and circle these score numbers on the score sheet.
3. By using your scores, determine those sentences (about 5–7 in total) which are the most central ones for the meaning of the article, by circling these sentences in the first column.

Note: Feel free to underscore or to make marks on the text sheets. Also, you are encouraged to make notes or comments about difficulties, problems or specifically interesting issues that arise during your work.
Please work through the articles in the order indicated on the Subject Form.

Figure 4: Instructions for Subjects in the Experiment.

3.3 Results

For every article, the mean sentence scores and the corresponding standard deviations were computed. Also, the Pearson r coefficients between all pairs of subjects were computed for every article and for all articles together. Additionally, a chi square test was performed for all pairs of subjects, both for two partitions of scores ("high relevance" (4 and 5) vs. "low relevance" (1, 2, 3)) and for the distinction "sentence selected for the abstract" vs. "not selected". These results, except for the mean values and standard deviations which will be shown graphically in the next section (for a comparison between machine and human results), are given in the tables 17 to 24 in the appendix C; ******* means "significant at the 0.01 level", ***** means "significant at the 0.05 level", whereas **n.s.** is short for "non significant".

Taking the sorted mean sentence scores as a basis, for every article, the highest ranked sentences were determined, using the same algorithm for threshold-shifting as implemented in the machine system. Tables 10 and 11 show the resulting sentence numbers, their ranks and scores. The corresponding abstracts are listed in the appendix C.

3.4 Some Comments by the Subjects

Only very few subjects made use of the possibility of marking or under-scoring text passages. However, a few comments about both general issues of abstracting and specific texts were made the more interesting of which are compiled together here:

	HS2	HS4	HS3	HS8	HS7	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
HS2		***	***	***	n.s.	***	***	***	***	***	***	***	***	***
HS4			***	***	n.s.	***	***	***	***	***	***	***	***	***
HS3				***	***	***	***	***	***	*	***	***	***	***
HS8					***	*	***	***	***	***	***	***	***	***
HS7						n.s.	n.s.	n.s.	n.s.	*	n.s.	***	n.s.	***
HS9							***	***	*	*	***	***	***	***
HS1								***	***	***	***	***	***	***
HS5									***	***	***	***	***	***
HS12										***	***	***	***	n.s.
HS11											***	***	***	***
HS13												***	***	*
HS10													n.s.	***
HS14														***

Table 9: Significance of sentence score correlation: All 6 articles

Article A			Article B			Article C		
number	rank	score	number	rank	score	number	rank	score
1	1	4.69	1	2	4.46	1	2	4.69
2	2	4.62	2	1	4.92	5	5	4.15
7	3	4.0	3	4	4.31	7	3	4.46
8	4	3.92	8	5	4.23	15	1	4.85
9	5	3.77	9	6	4.08	16	6	3.85
12	6	3.69	10	3	4.38	18	7	3.85
						21	4	4.38

Table 10: Optimal selected sentences, ranks and scores for articles A, B, and C.

Article D			Article E			Article F		
number	rank	score	number	rank	score	number	rank	score
1	1	4.77	1	1	4.85	1	1	4.77
3	5	3.54	7	2	4.23	2	4	3.85
8	6	3.46	10	5	3.77	3	2	4.15
9	2	3.85	12	6	3.69	6	3	4.15
10	4	3.62	13	7	3.69	9	6	3.77
16	3	3.77	14	3	4.0	13	5	3.85
			15	4	3.92			

Table 11: Optimal selected sentences, ranks and scores for articles D, E, and F.

Article	Mean Std. Dev.	Min. Std. Dev.	Max. Std. Dev.
A	0.976	0.63	1.38
D	1.09	0.599	1.39
F	0.964	0.439	1.41
B	0.858	0.277	1.25
C	0.927	0.376	1.26
E	0.96	0.555	1.13

Table 12: Mean, minimum and maximum standard deviations for sentence scores.

- "previous knowledge of the subject matter and/or specific interest helps in doing the abstracting task"
- "the first sentence seems to be important for most newspaper articles"
- "are the *most central* sentences always the *most important* ones for an *abstract*?"
- "sometimes several topics are merged; this makes abstracting difficult"
- "if there exist several threads, it is not clear how to decide which of these to select for an abstract" (article D)
- "anaphora resolution problems occur when creating *extracts as abstracts*"
- "very dense information is difficult to abstract" (article B)

3.5 Discussion

The results of the experiment show that there is a high correlation between subjects concerning the sentence scoring task. With the exception of three subjects (12, 13 and 14) there are also high correlations concerning the selected sentences for the abstract. If one looks at individual articles separately, however, due to the small number of data points the correlation values decrease and fall below or just above the significance level. Article C is an exception: here we get a very high inter-subject correlation for all subjects except for nr.9 and nr.12. An explanation might be that this article, dealing with a well-known issue in EC policy, was easy to understand and also well-structured.[35]

One strategy which seems to hold in general is that the first sentence of an article gets high scores and is selected very frequently by the subjects (see tables 10 and 11 and the relevant comment in subsection 3.4). However, there was no trend which would indicate any preference for sentences at the end of the articles. Only in article C, the last sentence would be included in the optimal abstract (see tables 10 and 11).

It also should be pointed out that relevant passages tend to consist of more than one sentence: in the abstracts for articles A and B, there occur two such groups, in the other four

[35] One subject commented on this, stating that "this article was easier than article D". However, another subject (nr.12, having less correlation to other subjects here, anyway) noted that "the whole article (C) seems a bit pointless."

	Article A		Article B		Article C	
	precision	recall	precision	recall	precision	recall
HS2	0.56	0.44	0.54	0.57	0.57	0.67
HS4	0.57	0.65	0.49	0.5	0.63	0.51
HS3	0.41	0.47	0.52	0.54	0.7	0.56
HS8	0.35	0.4	0.65	0.57	0.45	0.67
HS9	0.58	0.54	0.57	0.5	0.49	0.46
HS1	0.68	0.54	0.52	0.54	0.67	0.54
HS5	0.5	0.58	0.38	0.48	0.49	0.46
HS12	0.41	0.6	0.42	0.37	0.41	0.83
HS11	0.56	0.53	0.65	0.57	0.74	0.6
HS13	0.49	0.57	0.36	0.38	0.68	0.64
HS10	0.57	0.4	0.45	0.47	0.66	0.52
HS14	0.41	0.48	0.5	0.44	0.44	0.52
HS15	0.55	0.44	0.44	0.55	0.59	0.56
MEAN	0.51	0.51	0.5	0.5	0.58	0.58

Table 13: Mean Precision and Recall Across All Human Subjects (Articles A, B, C).

at least one, these groups consisting of three sentences on the average.[36] This is certainly an issue which should be considered when building automatic abstracting systems. To some extent, this amounts to using text-tiling methods which aim to detect coherent passages (and boundaries between them) in a text. Term and phrase clustering methods could also provide useful information for an abstracting system in that it may help to decide which sentences are "linked" together more or less strongly. Finally, the intricate problem of anaphora resolution would become less important in a system extracting passages instead of single sentences. One major drawback of this approach, however, "the price of coherence", would be the resulting decrease of precision which probably would not be compensated for by a corresponding increase of recall.

Since there are no significant differences in terms of variance between the two groups of articles (A,D,F vs. B,C,E; see table 12), the hypothesis that title and paragraph information is crucial for coherent results between different human abstractors cannot be confirmed — there is even a trend that the contrary might be true though this would have to be confirmed by a much larger sample size.[37]

We also computed the mean precision/recall values for each subject, compared to all the others; this was done for every article separately, and for all articles together. Also, the means for these inter-subject means were computed. The results are shown in tables 13 and 14. We find an overall precision/recall of 0.49 which corresponds to the results for machine method 3 (0.47/0.54), i.e., this method yields about the same precision/recall as an "average" human subject does. Precision/recall amongst the subjects are better than average for article

[36] The random chance of getting (at least) one three-sentence passage when forming an abstract of about 6 or 7 sentences from about 19 sentences — which are the values for the six test articles — is about 38%; getting such passages for at least 5 out of 6 articles (as it is the case in the experiment's results) has a random chance of only about 3%.

[37] Of course this does not mean that in a machine system such information would not be worth considering.

	Article D		Article E		Article F		All Articles	
	precision	recall	precision	recall	precision	recall	precision	recall
HS2	0.45	0.42	0.42	0.38	0.49	0.58	0.5	0.5
HS4	0.49	0.39	0.46	0.57	0.49	0.47	0.51	0.51
HS3	0.47	0.53	0.5	0.6	0.45	0.53	0.54	0.5
HS8	0.49	0.44	0.54	0.54	0.49	0.47	0.51	0.49
HS9	0.32	0.37	0.51	0.63	0.58	0.48	0.5	0.5
HS1	0.32	0.37	0.53	0.45	0.54	0.53	0.5	0.53
HS5	0.52	0.4	0.47	0.57	0.6	0.58	0.51	0.48
HS12	0.34	0.67	0.63	0.38	0.41	0.48	0.49	0.44
HS11	0.51	0.47	0.53	0.54	0.32	0.31	0.51	0.54
HS13	0.36	0.33	0.49	0.44	0.56	0.54	0.48	0.48
HS10	0.57	0.44	0.37	0.32	0.57	0.56	0.45	0.52
HS14	0.34	0.38	0.49	0.5	0.37	0.36	0.45	0.42
HS15	0.4	0.38	0.42	0.43	0.56	0.54	0.48	0.48
MEAN	0.43	0.43	0.49	0.49	0.5	0.5	0.49	0.49

Table 14: Mean Precision and Recall Across All Human Subjects (Articles D, E, F; All Articles).

C (0.58) and worse for article D (0.43) which confirms the subjects' intuitions mentioned earlier. In general, these results confirm the findings of Rath *et al.* (1961) about the low overlap of selected sentences between different human abstractors — in our study, the overlap was higher than in Rath's investigation which is probably due to the relatively high percentage of selected sentences from a document (about 30% in our case).

One of the major problems for the subjects turned out to be the difficulty of deciding which of the threads to follow to what extent in multi-topical articles. The same problem would be true for a high-level machine system which faces the additional difficulty in sorting out how many sub-topics a text consists of and how strong the (semantic) links are amongst them. In general, however, newspaper articles in this corpus tend to deal with one topic only; even if they appear to be multi-topical, these topics are somewhat closely related so that they could be considered as belonging together to some extent.

4 Comparison, Evaluation, and Discussion: Man vs. Machine

We have already looked at the precision/recall results for the various methods of the machine system in section 2; to summarise, it was demonstrated there that method 3, which does just the summation of all the content word weights in a sentence, had the highest precision/recall values, i.e., came closest to the average result of the subjects in the experiment (mean precision/recall over all 6 articles is 0.47/0.54).

We will now have a closer look at the sentence scores of the subjects in comparison to those of machine method 3, for all 6 articles (see figures 5 to 10).[38] Table 15 shows mean recall,

[38] For these graphs, the machine scores were transformed to the interval [0.0,5.0] and then shifted wrt. the difference of score-averages between human subjects and machine.

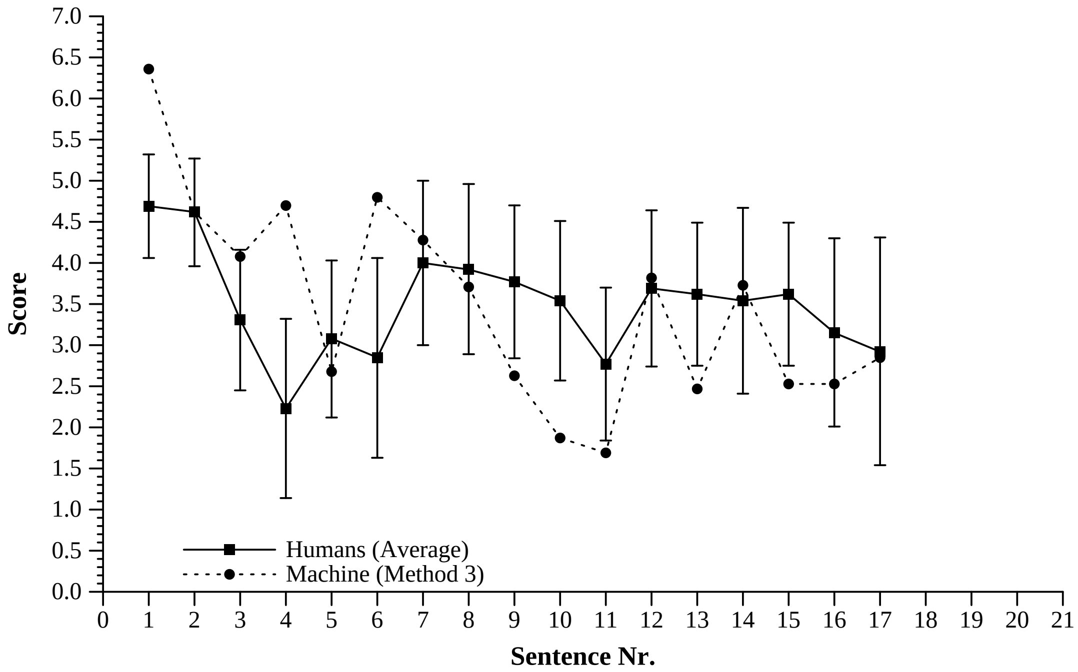


Figure 5: Article A: Sentence Scores: Humans (Average) vs. Machine (Method 3)

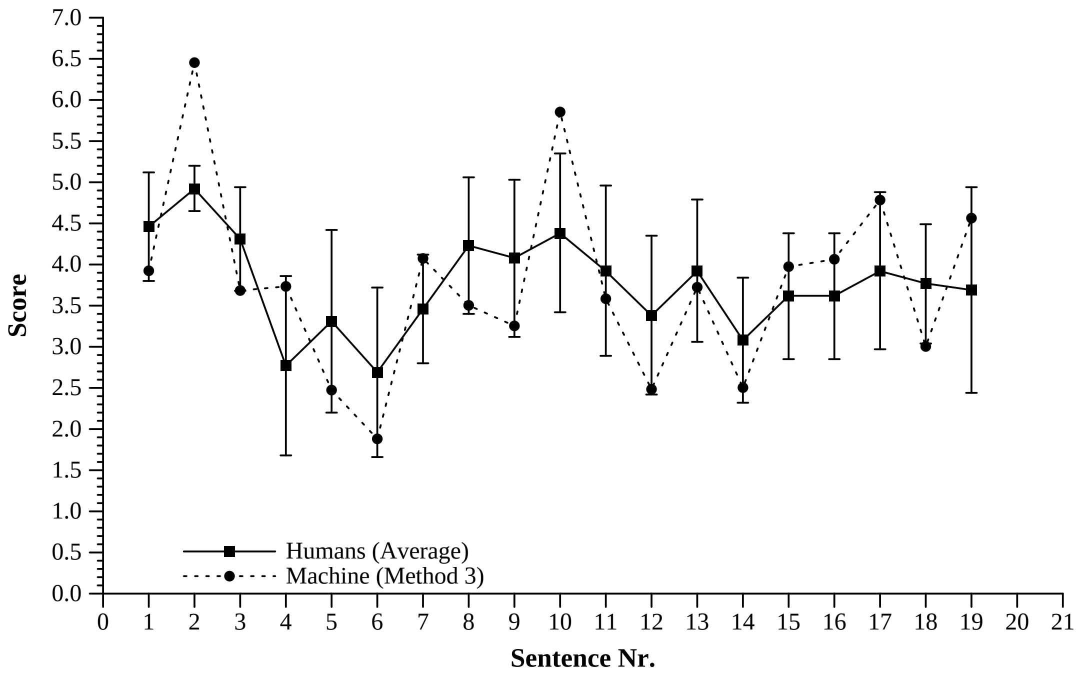


Figure 6: Article B: Sentence Scores: Humans (Average) vs. Machine (Method 3)

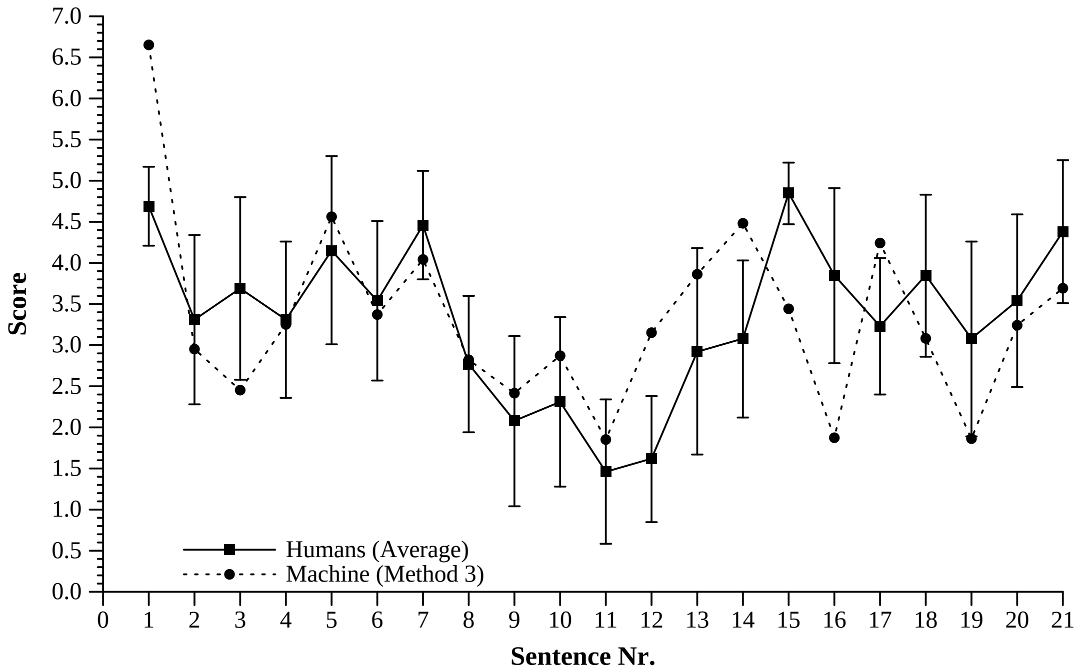


Figure 7: Article C: Sentence Scores: Humans (Average) vs. Machine (Method 3)

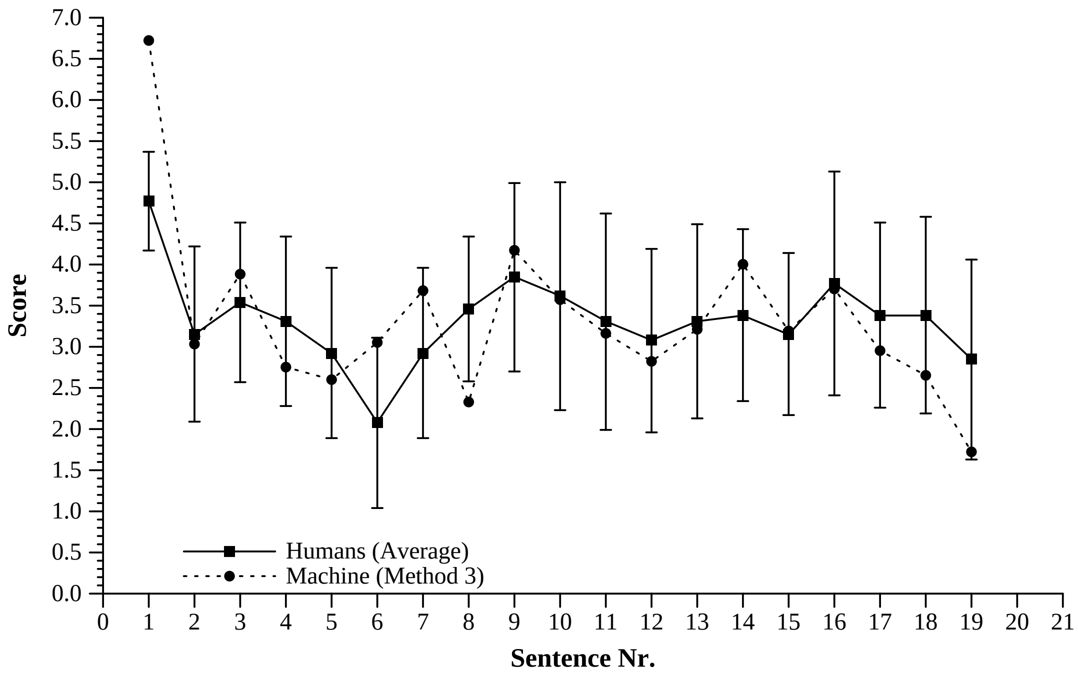


Figure 8: Article D: Sentence Scores: Humans (Average) vs. Machine (Method 3)

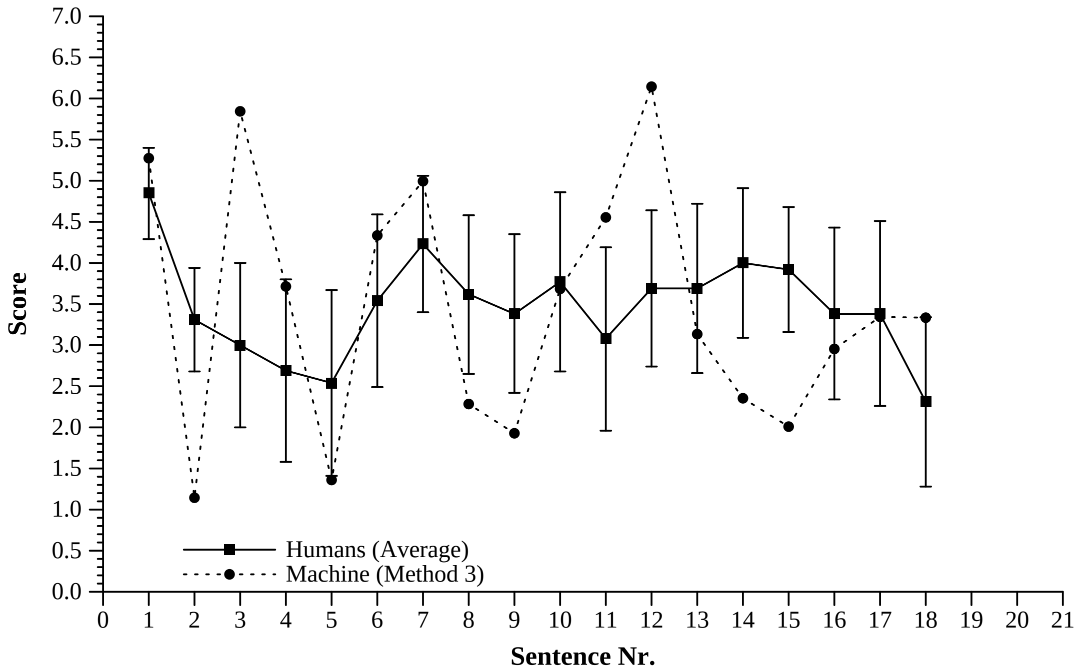


Figure 9: Article E: Sentence Scores: Humans (Average) vs. Machine (Method 3)

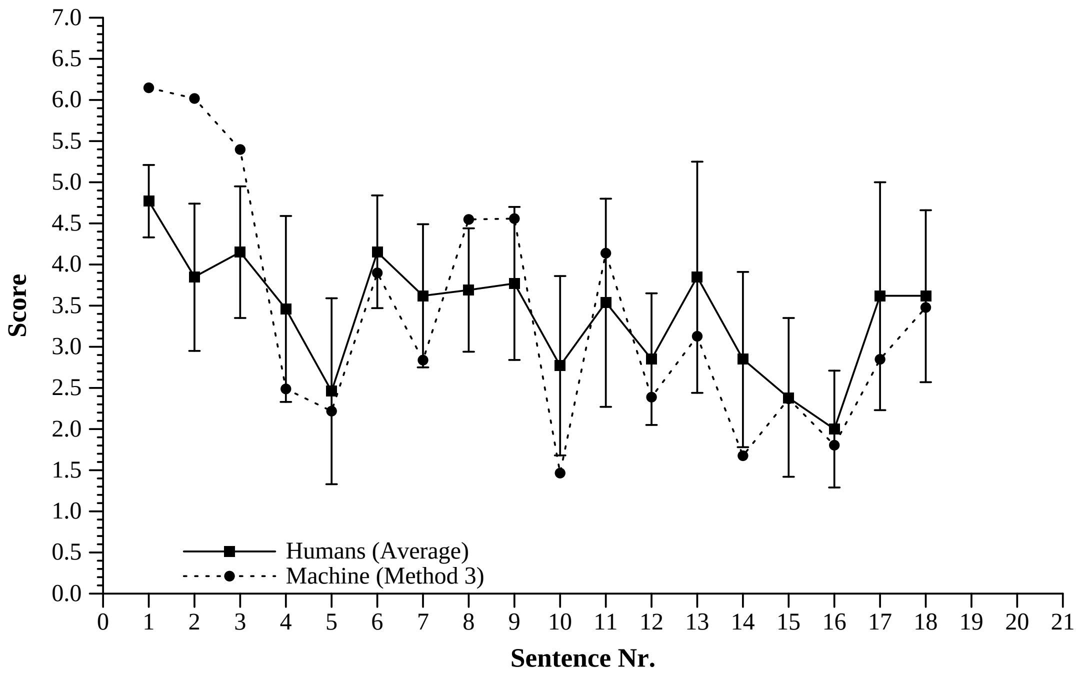


Figure 10: Article F: Sentence Scores: Humans (Average) vs. Machine (Method 3)

	Mean Precision	Mean Recall	Corr. At 0.01 Level	Corr. At 0.05 Level	No Sign. Corr.	Scores Within Std. Dev. (%)
Article A	0.48	0.6	2	0	11	52.9 (9/17)
Article B	0.42	0.55	3	4	6	84.2 (16/19)
Article C	0.45	0.55	0	9	4	61.9 (13/21)
Article D	0.46	0.49	2	4	7	89.5 (17/19)
Article E	0.47	0.44	0	2	11	55.6 (10/18)
Article F	0.52	0.62	4	4	5	66.7 (12/18)
Total	0.47	0.54	12	1	0	68.8 (77/112)
Without 1st Sent.	0.4	0.46	6	6	1	70.8 (75/106)

Table 15: Precision, Recall, Correlation and Closeness of Sentence Scores for Machine Method 3, Compared to Human Results.

precision, correlations to human results, and the percentage of sentences where the machine results are within the standard deviations of the subjects' mean scores.[39] It can be seen that for all of these articles, the sentence scores for method 3 stay within the standard deviation interval of the average subjects' scores for more than the half of the sentences; for articles B and D this is true for over 80% of the sentences. Globally, more than two thirds of the sentences fall into this interval. The weakest results are those for articles A and E, in terms of correlation and closeness of sentence scores.

The precise values for all Pearson r coefficients, their significance, and all precision/recall values for all machine methods across all subjects can be found in appendix D.[40] Also, the significance results of the chi-square tests for both sentence scores and selected sentences are presented there. It has to be mentioned that since the first sentence in an article is very likely to be included in an abstract (for both human subjects and the machine, due to its weighting factors), this fact in itself yields a (sometimes much) better correlation; when omitting the first sentences in the evaluation, the correlation and chi square values decrease and hence we get less significant results here (see also the tables in appendix D).

When looking at some individual sentences where the machine results differ a great deal from the human subjects' mean scores, we can identify a couple of possible reasons for these differences:

- *lack of term clustering:* An example is sentence 9 in article A which is weighted too low in the machine system:

    > They say the public right to cross land in Scotland is often wrongly assumed to be a general right [9].

    This sentence contains words or phrases which are quite closely related to title words like *land* to *countryside*, or *cross land* to *trespass*. In a clustering system, these semantic

[39] Since the subjects' scores are discrete values and the machine scores are in a continuous interval, these values concerning sentence scores within the interval of standard deviations are to be regarded as being indicative only.

[40] The machine results were produced using the default parameters as shown in table 4.

links could be identified and hence a higher word weight of these items would result in a considerably higher sentence weight.

- *short sentences being important tend to have too small weights*: This is a direct consequence of method 3 which sums over all content word weights in a sentence; it would have to be investigated if methods which *favour* — instead of discourage, as was suggested by Kupiec *et al.* (1995) — short sentences (to some extent) could improve the results. If such a sentence contains words appearing in a prominent section of an article, e.g., in the title or in the first sentence, it might be worth considering to (further) increase the respective word weights. An example of such an under-estimated sentence is sentence 8 in article D:

  > The risk, he said, is that the Whitewater controversy will paralyse the Presidency [8].

  Here, the word *Whitewater* appears both in the title and in the first sentence and is certainly the most important index word of the article. Although it is top ranked, one could argue for a further weight increase here and in similar cases.

- *background (world) knowledge seems to be required:* There are examples, such as sentence 3 in article E, where it is hard to imagine how a system solely employing statistical methods could figure out its, in this case, relatively low importance:

  > The first VC-10 load of troops, an advance party from the 1st Bn The Duke of Wellington's Regiment, left RAF Brize Norton, Oxon, within an hour of Mr Malcolm Rifkind, the Defence Secretary, announcing to the Commons that the Government had decided to send the extra forces to support the UN's request for 10,650 [3].

  This sentence is long *and* contains quite a few high ranked words, such as *troops*, *UN*, *Defence*, *announce*, and *Government*. Still, subjects decide this not to be a very important sentence, it is not included in the "optimal abstract" (see table 11). However, with machine method 3 it gets rank two.

In many of the sentences with high deviant scores, however, the "trend" of the machine results is right, i.e., it sort of "exaggerates" positive or negative peaks in the human average graph; since the latter is "naturally smoothed" (via the average computation), these results are not very surprising.

5 Further Suggestions and Conclusions

5.1 Possible Improvements of the System

5.1.1 Computational Issues

Most of the computational suggestions listed here have already been mentioned in various sections of this dissertation; therefore I will expand here only on those which have not been discussed earlier:

- introducing term and phrase clustering methods
- employing methods of text-tiling, possibly using paragraph-boundary information here
- exclude more low frequency words (e.g., below 10 occurences in the corpus, for a discussion see e.g. Salton & McGill (1983: 62)); this reduces the size of the corpus word list file significantly
- store the idf-values directly in the corpus word list file, instead of computing them in the program
- reading the corpus word list file only once when processing multiple articles
- carefully select a hash-function which proves to be most efficient for the task of handling text-strings
- possibly distinguish between upper-case and lower-case words
- using key or clue words and phrases for increasing or decreasing sentence weights
- possibly using word weight factors for words occuring in the first sentence as it is done for words in the title
- using global word frequency information: In our system, only the (standard) inverse document frequency is used for word weight computation; however, it could make sense to distinguish, say, between two words both occuring in 200 documents but having a different global word frequency. This would, e.g., amount to an additional weighting factor which favours words having a high "corpus-frequency/words-in-document-frequency" ratio. A formula for this weighting factor f might be

$$f = \frac{\alpha * N + \mathrm{f}corpus}{\alpha * N + \mathrm{f}art} \tag{7}$$

  where N is the number of documents in the corpus, $\mathrm{f}corpus$ is the corpus frequency and $\mathrm{f}art$ the word-in-document frequency of a word, i.e., in how many different articles it occurs. The factor α, when set judiciously, ensures that words having a (very) low corpus frequency but a high $\mathrm{f}corpus/\mathrm{f}abstr$ ratio are not given too high weighting factors.

If a sufficiently large set of abstracts produced by human abstractors were given, the system could be automatically trained to optimise its various parameters in the following way:

1. compute the mean sentence scores for all sentences in all abstracts
2. split the abstracts in a training and a testing set
3. optimise the system's parameters by minimising the difference of its output vectors to those of the training set
4. compute the system's performance with the testing set

By means of a multiple regression analysis, the relative influence of the various system parameters on the total results could be determined; these results could be used for a redesign of the original system.

5.1.2 User Interface Issues

So far, we have looked at the system as a “stand-alone” implementation only, i.e., without its interface to the user. If it were to be set in an off-line, non-interactive environment, this would be entirely justified, but in practice it is realistic to assume that automatic abstracting systems will be offered as additional features to already existing (standard) IR systems which nowadays allow an increasing amount of user-interaction.

A very obvious point was already mentioned in section 1.3.1: The abstracting system could use the information from the user query as a help to determine the relevant sentences from those documents it retrieves. However, it would have to be investigated, how “strictly” this should be done, i.e., how much weight-increase those words appearing in the user query should get.[41] A possibility would be to highlight the passages containing words from the user query to focus her attraction on these. Presenting such abstracts, and maybe also a list of top ranked keywords might also help the user to reformulate her query in a more precise way and hence being useful in a general IR paradigm, as well.

Concerning the output, a design should be chosen where it is easy for the user to determine to which parts of the document the extracted sentences belong to. This could be done via a small graphics, or at least by indicating the lengths of the gaps between the extracted passages. If, e.g., unresolved anaphora appear in any of these passages, the user would be able to ask the system for some additional previous context.

5.2 Suggestions for Additional Evaluations and Experiments

It would be worth investigating which factors are the most “important” ones for humans in doing text abstracting. Possible issues one could look at are, e.g., correlations to sentence length, position within the article as a whole and/or within paragraphs, correlation to different content or clue words. These results might help in improving the machine system by relying more on the techniques used by humans than this has been the case so far.

Another task would be to sort out which articles and which parts of them were particularly easy or difficult to abstract for humans: this might, on the one hand, give some clues about possible problems a machine system might face in certain texts, and, on the other hand (and as a very long term goal), yield a kind of a catalogue of properties which an “easy-to-understand” text should have. Of course, one could argue that there is not much of a point to give directions to human writers which tell them how to write in order to be easily understood (and “abstracted”) by machines, but if the matter of text-understanding applies to other humans, i.e., the readers of their texts, as well, it might convince some people to adhere to certain “writing rules”.

If one had built a fairly satisfying automatic abstracting system, performing a kind of a “Turing test” would certainly be an interesting issue.[42] A possible set-out might involve a couple of machine- and human-made abstracts which are given to people who then have to

[41] In fact, by using a title weight factor, our system could be seen as using the title as a query, though, not exclusively.

[42] Although, as we have seen even for our small system with quite a few drawbacks, it is not too difficult to get as good as abstracting results with that as with using a human abstractor.

score their quality and rank them, not knowing that some of these abstracts were produced by a machine system. If it turns out that the machine results are not significantly classified as being of worse quality than those of the human abstractors, one could argue that the machine-system has passed this (simple) test.

5.3 General Issues and Final Remarks

When an automatic text abstracting system is to be embedded into an existing IR system, the only additional task which has to be performed would be the calculation of sentence weights, since the word weights (tf*idf) are to be computed anyway. These weights or just the rank order of the sentences would then be stored together with the document, requiring almost no additional disk space (one integer per sentence). When the user chooses the abstracting option, the system just has to retrieve the top ranked n sentences and output these in text order. If the user increases or decreases the required length of the abstract, sentences have to be inserted or deleted at the appropriate location. Highlighting methods, as mentioned earlier, can help the user to focus her attention to her specific interests, according to the user query she gave to the system initially.

Finally, the issue of low-level vs. high-level systems should be discussed shortly. The work described here has shown quite clearly that good abstracting results can be achieved by means of using statistical methods only. Of course, there are several restrictions which were already mentioned, e.g.

- the documents (articles) are fairly small, and
- they are mainly mono-topical
- certain domains, specifically those containing dense information, like articles about finance/business, are difficult to deal with
- certain important problems, like coherence, anaphora resolution, dealing with various ambiguities are not solved or even being attempted to be solved

However, taking the high-level approach, either as add-on or as an alternative entirely, would face one with the yet unresolved and much greater problem of the design of large knowledge bases for a broad range of domains, irrespective of the fact if one chose to do (fast) pattern matching or rather (deep) parsing and semantic analysis. On the other hand, it is quite clear and obvious that a merely statistically based system never gets to some sort of “text understanding”. But maybe some people are worried about this for wrong reasons: do we have to assume that text abstracting has to involve text understanding, just by analogy to how we think humans perform this task?

It will remain true for some time that high-level abstracting systems are useful for research purposes or for very restricted domains only, whereas building an abstracting system which has to cover a general and broad range of domains requires relying on low-level statistical techniques, together with some additional fine-tuning using heuristic parameters.

6 Summary

In this paper, we have shown that it is possible to implement a system mainly based on word frequency statistics which generates abstracts from newspaper articles by computing word and sentence weights and extracting the top ranked sentences.

It was demonstrated that the resulting abstracts have the same quality in terms of precision/recall as the abstracts created by human subjects in an experiment.

Further, there is a high overall correlation concerning the sentence scores both between the machine system and the human subjects, and between the human subjects themselves.

7 References

Bookstein, A., Klein, S.T., Raita, T. (1995): Detecting Content-Bearing Words by Serial Clustering — Extended Abstract. In: Proceedings of the 18th ACM-SIGIR Conference (1995), pp.319–327

Callan, J.P. (1994): Passage-Level Evidence in Document Retrieval. In: Proceedings of the 17th ACM-SIGIR Conference (1994), pp.302-310

Church, K.W. (1995): One Term or Two? In: Proceedings of the 18th ACM-SIGIR Conference (1995), pp. 310–318

DeJong, G.F. (1979): Skimming Stories in Real Time: An Experiment in Integrated Understanding. PhD dissertation, Yale University

Edmundson, H.P. (1969): New Methods in Automatic Extracting. In: Journal of the ACM, 16 (2), pp.264–285

Goldfarb, C.S. (1990): The SGML Handbook. Oxford University Press: New York

Guilford, J.P., Fruchter, B. (1978): Fundamental Statistics in Psychology and Education. Sixth Edition. McGraw-Hill: New York etc.

Hearst, M.A., Plaunt, C. (1993): Subtopic Structuring for Full-Length Document Access. In: Proceedings of the 16th ACM-SIGIR Conference (1993), pp. 59–68

Hobbs, J.R., Appelt, D.E., Bear, J.S., Israel, D.J., Tyson, W.M. (1992): FASTUS: A System for Extracting Information from Natural Language Text. SRI International, Technical Note 519, Menlo Park, CA

Jacobs, P.S., Rau, L.F. (1990): SCISOR: Extracting Information from On-line News. In: Communications of the ACM, 33 (11), pp.88–97

Kupiec, J., Pedersen, J., Chen, F. (1995): A Trainable Document Summarizer. In: Proceedings of the 18th ACM-SIGIR Conference (1995), pp.68-73

Lehnert, W.G. (1980): The Role of Scripts in Understanding. In: Metzing (1980), pp.79–95

Luhn, H.P. (1959): The automatic creation of literature abstracts. In: IBM Journal for Research and Development, 2 (2), pp.159–165

Mann, W.C., Matthiessen, C.M.I.M., Thompson, S.A. (1989): Rhetorical Structure Theory and Text Analysis. ISI Research Report ISI/RR-89-242, University of Southern California, Marina del Rey, CA

Mauldin, M.L. (1989): Information Retrieval by Text Skimming. CMU-CS-89-193, Carnegie Mellon University, Pittsburgh, PA

Metzing, D. (ed.) (1980): Frame Conceptions and Text Understanding. De Gruyter: Berlin

Miike, S., Itoh, E., Ono, K., Sumita, K. (1994): A Full-Text Retrieval System with a Dynamic Abstract Generation Function. In: Proceedings of the 17th ACM-SIGIR Conference (1994), pp.152–161

Minsky, M. (1980): A Framework for Representing Knowledge. In: Metzing (1980), pp.1–25

Mittendorf, E., Schäuble, P. (1994): Document and Passage Retrieval Based on Hidden Markov Models. In: Proceedings of the 17th ACM-SIGIR Conference (1994), pp.318-327

Paice, C.D., Jones, P.A. (1993): The Identification of Important Concepts in Highly Structured Technical Papers. In: Proceedings of the 16th ACM-SIGIR Conference (1993), pp.69–78

Rath, G.J., Resnick, A., Savage, T.R. (1961): The formation of abstracts by the selection of sentences. In: American Documentation, 12 (2), pp.139–143

Rijsbergen, C.J. van (1979): Information Retrieval. Butterworths: London. Second Edition

Salton, G. (1989): Automatic Text Processing. Addison-Wesley: Reading, Mass. etc.

Salton, G., Allan, J., Buckley, C. (1993): Approaches to Passage Retrieval in Full Text Information Systems. TR 93-1334 (1993), Cornell University, Ithaca, NY

Salton, G., Allan, J., Buckley, C., Singhal, A. (1994): Automatic Analysis, Theme Generation, and Summarization of Machine-Readable Texts. In: Science, 264, pp.1421–1426

Salton, G., Buckley, C. (1990): Flexible Text Matching for Information Retrieval. TR 90-1158 (1990), Cornell University, Ithaca, NY

Salton, G., McGill, M.J. (1983): Introduction to Modern Information Retrieval. McGraw-Hill: Tokyo etc.

Schank, R.C., Goldman, N.M., Rieger, C.J., Riesbeck, C.K. (1975): Conceptual Information Processing. North-Holland: Amsterdam. Fundamental Studies in Computer Science, Vol. 3

Schildt, H. (1995): C++: The Complete Reference. Second Edition. McGraw-Hill: Berkeley etc.

Sundheim, B. (ed.) (1992): Proceedings of the Fourth Message Understanding Conference (MUC-4), McLean, Virginia. Morgan Kaufmann Publ.: San Mateo, CA

Wilkinson, R. (1994): Effective Retrieval of Structured Documents. In: Proceedings of the 17th ACM-SIGIR Conference (1994), pp.311-317

Acknowledgements

My thanks go to my supervisors for their valuable help and advice throughout this project. Also, I wish to thank my parents for their ongoing support in all respects; this work is dedicated to them.

A The Sample Newspaper Texts

Article A: Countryside Alliance Fights Trespass Law.

By Charles Clover, Environment Editor.

AN unprecedented alliance of 115 countryside, wildlife and sporting organisations has written to Mr Ian Lang, the Scottish Secretary, asking him to scrap proposals making trespass a criminal offence in Scotland [1].

They express concern that parts of the Criminal Justice and Public Order Bill, which gives the police new powers to control the disturbance of hunting and grouse shooting, could be used by landowners to "intimidate" people using the outdoors for recreation [2].

The letter says the new offence of aggravated trespass "potentially affects every man, woman or child who wants to enjoy Scotland's countryside" [3].

It is signed by Mr Bill Murray, chairman of the Scottish Countryside Activities Council, Mr Fred Nelson, chairman of the Scottish Sports Association, and Mr Seaton Baxter, chairman of Scottish Wildlife and Countryside Link [4].

The groups say intimidation could take the form of restrictive notices, verbal confrontations and the threat of the involvement of police officers to investigate alleged criminal offences [5].

Mr Lang has already been urged by Scotland's police constables and by Magnus Magnusson, chairman of Scottish Natural Heritage, the country recreation watchdog, to scrap the proposed new offence in Scotland [6].

The new alliance of outdoor groups, which includes ramblers, mountaineers, cyclists, skiers and wildlife enthusiasts, says the Government has taken no account of the different tradition and legal basis for public access in Scotland and has made no attempt to consult with any of their organisations [7]. "There appears to be no recognition of the fact that virtually everyone going for a walk or run in Scotland's countryside is, technically speaking, exposed to accusations of being a trespasser," say the chairmen [8].

They say the public right to cross land in Scotland is often wrongly assumed to be a general right [9]. It is instead almost entirely based on assertion by the public of an accustomed right [10].

This, they say, is "a traditional right which many of the members we represent cherish deeply" [11].

The number of footpaths, tracks and wild areas where the public has secured a legal right to roam are "minute" when compared to the area of land to which the public presently enjoys access in Scotland - or with the official rights of way network in England and Wales [12].

The introduction of a crime of aggravated trespass would, they argue, lead to a "confrontational" situation with landowners [13].

They note that the provisions in the Criminal Justice Bill which seek to control "raves" - or mass parties - apply only to England and Wales and suggest the same should apply to aggravated trespass [14].

They say the need for measures to control the disruption of hunting and grouse shooting in Scotland has yet to be demonstrated [15]. Even if disruption were to take place it would be relatively easy for the Government to bring in an amendment to a future Criminal Justice Bill [16].

They add that the present Bill makes a nonsense of a review of access to the outdoors being conducted by Scottish Natural Heritage [17].

Article B: Lyell 'Did Not Report Heseltine Concern'.

By Sean O'Neill.

THE ROLE of Sir Nicholas Lyell, the Attorney General, in the arms-to-Iraq affair was called into question again yesterday, shortly before he appears at the Scott inquiry [1].

Mr Alan Moses, QC, leading prosecution counsel in the Matrix Churchill trial, told the inquiry that Sir Nicholas did not inform him that Mr Michael Heseltine, President of the Board of Trade, had grave reservations about signing a public interest immunity certificate [2].

The trial of three former executives of Matrix Churchill, charged with illegally exporting arms-making machinery to Iraq, collapsed in November 1992, after certificates were overturned by the judge, releasing to the defence Government papers showing ministerial knowledge of the trade [3].

Mr Moses told Lord Justice Scott that the use of such certificates in the Matrix Churchill case, which has been defended by Sir Nicholas, had undermined public confidence in justice [4]. Such evidence from the Government's lawyer has further undermined the Attorney General, who is to give evidence at the inquiry tomorrow [5].

He is certain to face a gruelling interrogation about the quality of his advice to ministers [6].

On his second day before the inquiry, Mr Moses said he would have "wanted to know" about Mr Heseltine's views, but because he had not been told of them he felt he had not represented the minister's position properly at the trial [7].

Last month the inquiry was told that Mr Heseltine had refused to sign a gagging order to withhold Whitehall papers from the defence in the case because he feared he would be taking part in a "cover-up" [8].

Sir Nicholas told Mr Heseltine that despite his concerns it was his ministerial "duty" to sign [9]. Mr Heseltine agreed to sign a much amended certificate after Sir Nicholas assured him that the "unusual wording" and "limited scope" of the order would be drawn to the attention of the judge by counsel, yet three days later, on Sept 10 1992, Sir Nicholas met Mr Moses but did not mention the issue [10]. Mr Moses had further contacts with the Attorney's office and with the Treasury Solicitor's Department, but no one instructed him that Mr Heseltine wanted his dissenting view "flagged up" to the court [11].

On Sept 30, Mr Moses went to court and argued to uphold the certificate [12]. He told Judge Smedley, the trial judge, not that Mr Heseltine had been a reluctant signatory, but that the four ministers who had signed gagging orders were of one mind - that disclosure of documents would damage the public interest [13].

Mr Moses admitted: "My submission was not a sufficient explanation of the President's views as I now understand them [14]."

Mr Moses's evidence to the inquiry raised doubts about a statement issued by Sir Nicholas after Mr Heseltine's appearance at the inquiry [15].

Sir Nicholas had said that Mr Heseltine's certificate had been "expressly amended after consultation with the Attorney General and leading counsel for the prosecution" [16]. Mr Moses told the inquiry that he had received a copy of Mr Heseltine's amended certificate but he was not told what changes Mr Heseltine had made, or their significance [17].

"What I did not take on board and did not appreciate was that the President had expressed positive views as a matter of independent judgment about the individual documents," said Mr Moses [18].

He added: "I can well see a minister would be very ill-advised to sign a certificate unless he was quite satisfied as to the reasoning of the prosecution for fear, as I think Mr Heseltine feared, that he would be accused of withholding documents ... which a judge has ordered to be disclosed [19]."

Article C: Major Is Ready For Long Fight With EC.

By George Jones and Christopher Lockwood.

Britain's rift with the rest of Europe deepened last night after EC foreign ministers failed to break the deadlock over voting rights and Mr John Major, adopting a strident anti-Brussels stance, signalled that the Government was digging in for a long battle [1].

The Prime Minister will now seek to use the issue to the Conservatives' advantage in the forthcoming elections for the European parliament [2].

However his confrontational tone heightened fears of a new Tory split over Europe [3]. Pro-European Tory MPs accused the Prime Minister of bowing to the views of the party's Euro-sceptics and claimed that Mr Douglas Hurd, the Foreign Secretary, was unhappy with Mr Major's tough stand [4].

Yesterday's talks in Brussels broke up when it became clear there was no compromise acceptable to Britain and Spain - the only other country opposing an increase in the number of votes needed to block EC legislation [5].

Another effort will be made to broker a settlement when EC foreign ministers meet in Greece at the weekend, but positions on both sides appear to be hardening [6].

Britain's European partners said that failure to solve the voting issue within the next few days could derail plans to enlarge the Community from 12 members to 16 with the inclusion of Austria, Sweden, Norway and Finland next year [7].

While Mr Hurd emerged from yesterday's talks emphasising that Britain was "not inflexible", Mr Major was in uncompromising mood in the House of Commons [8].

Euro-sceptics cheered when he vowed to "fight Britain's corner hard" [9]. He said "phoney threats" to delay enlargement of the EC would not sway him [10].

He also took an electioneering swipe at the Labour Party [11]. Claiming Labour would sign away "competitiveness and money", he described Mr John Smith, the Labour leader, as: "The man who likes to say 'yes' in Europe - Monsieur Oui, the poodle of Brussels [12]."

Delighted Tory Euro-sceptics claimed that Mr Major, faced with the prospect of a Conservative "civil war" or a clash with Europe, had chosen to battle with the EC [13].

But there was dismay among the party's European wing, who fear enlargement of the EC is threatened and that Mr Major is jeopardising compromise within the party over the issue [14].

The voting row centres on whether the number of votes in the EC's ruling Council of Ministers needed to block legislation should be increased from 23 to 27 when the four new countries join [15].

There remains strong opposition within the Cabinet to giving way on 27 [16]. Mr Kenneth Clarke, the Chancellor, said during a visit to Sweden that London felt the voting issue should be uncoupled from the enlargement process and deferred until 1996, when EC leaders are to hold an inter-governmental conference [17].

Emerging from the Brussels talks, Mr Hurd said he had rejected a compromise proposal, which would allow for a delay of one or two months in cases where there were 23 votes against a proposal, but not the full 27 [18].

"In our view, this does not give adequate protection to our interests," he said [19]. The compromise would have included a non-binding declaration that the EC would try to avoid imposing decisions blocked by 23 votes [20].

Britain is demanding a protocol to the enlargement treaty which would commit the EC not to over-ride a vote of 23 within a still unspecified time limit [21].

Article D: International - Senate To Stage Public Inquiry On Whitewater.

By Maurice Weaver in Washington.

WASHINGTON began preparing yesterday for one of the biggest Congressional spectaculars since the Iran-Contra inquiry following a Senate decision to hold public hearings on President Clinton's Whitewater property dealings [1].

Leaders of the minority Republican Party, having bulldozed Democratic objections aside, hope the "show" will start in the spring [2].

But first they must agree a timetable with the Democrats who, despite having caved in, remain deeply uneasy at the prospect of seeing the President's pre- election record of private financial manoeuvres subjected to public scrutiny [3].

The television and press coverage will be unwelcome and possibly damaging at a crucial stage in Mr Clinton's legislative programme [4].

The President, faced with a fait accompli, has said diplomatically that Congress must "do whatever it is they think is the right thing to do" [5].

But one of his closest political advisers, Mr James Carville, gave a more candid insight into White House feelings yesterday, when he grumbled that the affair was turning into "a zoo" [6].

Another White House insider, Presidential Counsellor David Gergen, expressed dismay at the "overheated atmosphere" in Washington and the vehemence of Mr Clinton's critics, which he attributed largely to his ambitious reform programme [7]. The risk, he said, is that the Whitewater controversy will paralyse the Presidency [8].

Even as the Clinton team rallied to the banner, however, new information was leaking out about the Clintons' family finances which, while not suggesting any improprieties, does indicate how the couple benefited from the "cronyism" in their native state of Arkansas [9].

Mrs Hillary Clinton, the New York Times disclosed, made nearly #660,000 by trading in cattle futures in the late 1970s on the advice of a lawyer friend who worked for Tyson Foods, one of Arkansas's biggest companies [10].

The revelation is considered significant because it shows how a couple from relatively humble beginnings were able, by virtue of knowing the right people, to attain the financial security that would help them scale the social and political heights [11].

The Clintons arrived in Washington as the ultimate meritocrats, the baby- boomers who made it to the very pinnacle on their own abilities [12]. How much damage will be done to that reputation by the disclosures that will come spilling out at public Congressional hearings is the key question [13].

Yesterday, the leader of the Senate's Democratic majority, Senator George Mitchell, and his Republican opposite number, Senator Robert Dole, were discussing with their inner circles what an "appropriate" timetable for the hearings will be [14].

The Democrats initially argued that a Congressional inquiry would hinder the investigation being carried out by Mr Robert Fiske, the special counsel appointed by the Justice Department [15].

But with the collapse of unanimity within the party as more members concluded that full public disclosure of the Whitewater facts was preferable to perceived secretiveness, the leadership was forced to give way to Republican demands [16].

Senator Dole is now calling for the hearings to begin as soon as possible, and his representatives speak unofficially of a spring start [17].

It would, theoretically, be possible for Congress to call Mr Clinton to testify [18]. But constitutionalists consider that highly unlikely [19].

Article E: International - Troops Leave For Bosnia Within Hour Of Defence Secretary Announcing Government Decision.

British forces will help UN monitor delicate ceasefire.

By Peter Almond, Defence Correspondent, and Robert Fox in Zagreb.

THE British troops who began arriving in Split last night will help the United Nations monitor a delicate Croat-Muslim ceasefire in Bosnia [1]. They will initially be deployed there for four months [2].

The first VC-10 load of troops, an advance party from the 1st Bn The Duke of Wellington's Regiment, left RAF Brize Norton, Oxon, within an hour of Mr Malcolm Rifkind, the Defence Secretary, announcing to the Commons that the Government had decided to send the extra forces to support the UN's request for 10,650 [3].

At the same time, six C-130 Hercules carrying Land Rovers left RAF Lyneham for Split, and three Navy auxiliary ships began loading 109 Saxon armoured troop carriers and other vehicles at Marchwood, Southampton [4].

They are expected to sail this morning and arrive in Split on March 20 [5]. By then, 792 soldiers, including engineers and logistics troops, along with 113 men with 15 Scimitar tracked reconnaissance vehicles from D Squadron of The Light Dragoons at Hohne, Germany, will have been deployed to Bosnia to join 2,300 already there [6].

Explaining the Government's change of mind about extra troops, Mr Rifkind told the Commons that the ceasefires had brought new opportunities and responsibilities that were greatly stretching the ability of the Coldstream Guards battalion group to fulfil its UN mission [7].

"The Coldstream Guards, whose mission hitherto had been to support humanitarian aid convoys, have found themselves with a major peacekeeping task on their doorstep," he said [8].

"It has become clear that the effort involved, while tolerable for a time, is unsustainable beyond the short term with their current manpower [9]. Although the UK contribution to the region is already a large one, a further UK contribution at this stage as part of a co-ordinated international effort would help to make the difference between success and failure for the ceasefires [10]."

The Duke of Wellington's will provide Lt-Gen Sir Michael Rose, the UN Commander in Bosnia, with his most powerful new body to help cement the Croat-Muslim ceasefire [11].

The troops will work to the complex ceasefire map drawn up by a joint commission of Bosnian army and Croat HVO militia commanders under the chairmanship of Brig John Reith, Commander of the new UN Sector SouthWest in Bosnia [12].

British officials say the map shows a ceasefire to be policed along 125 miles of front lines between the two sides, some of which cross high mountain ridges [13].

The soldiers are needed to patrol no man's land, help take weapons to collection sites and man checkpoints and observation posts [14].

Next week, the UN soldiers will begin overseeing the disengagement of forces [15].

Surprisingly, only a few miles of the front line are currently under dispute - part of the perimeter of Mostar airfield and a heavily booby-trapped municipal office in the middle of Vitez [16].

But UN commanders fear that any part of the front line could flare into conflict at any time, as the war is a hotchpotch of village rivalries and local feuds [17].

Jim Muir in Sarajevo writes: Gen Rose welcomed the announcement from London and other indications that he is likely to get the extra troops he needs [18].

Article F: 'Thatcher Circle' In Pergau Row.

By Robert Shrimsley and Anthony Looch.

The "close circle" around Lady Thatcher was criticised yesterday by Sir David Steel, the Liberal Democrats' foreign affairs spokesman, for involvement in the Pergau Dam project [1].

Speaking in an acrimonious Commons debate on the #234 million of British aid for the dam, Sir David said: "In the course of batting for Britain, which our former Prime Minister did extremely well, her close circle often seems to have been involved [2].

"In this case Sir Tim Bell, well known as her PR adviser, is also adviser to the Malaysian Prime Minister and to Tam Sri Armugam, who controls GEC Malaysia, heavily involved in several of the contracts under the deal [3].

"One other person who helped broker parts of the deal is Mr Steve Tipping, a business associate of Mr Mark Thatcher and indeed best man at his wedding [4].

"Tim Bell told the Sunday Times when asked to clarify Tipping's exact role: 'What he does for a living is introduce people to each other [5].' "

Opening the debate, Sir David said the only possible explanation for the funding of the project was the Government's "eagerness to lubricate the trade channels with Malaysia, especially the arms trade" [6].

He said Malaysia, which last year received #20.4 million in aid - the 12th largest donation from Britain - was far from poor [7]. Seeking to establish a link between aid and arms sales, he pointed out that Oman, whose GNP per head is higher than Portugal's had had its aid from Britain doubled since 1979 "and it currently ranks third largest in the list of purchasers of arms from Britain [8]."

Referring to an "unhealthy dominance" of Tory Party benefactors among firms in the Pergau project, Sir David said: "Companies linked by such donations have been the main beneficiaries to the tune of 42 per cent of the Aid and Trade Provision [9]." He cited Trafalgar House, Balfour Beatty, GEC and Biwater [10].

Mr Alistair Goodlad, a Foreign Office minister, ridiculed Sir David's figures, adding: "It is ridiculous to suggest that a country where there is a defence sales relationship should, because of that relationship, be ineligible for aid for trade provision [11]."

He said: "If we had not carried through our commitment our credibility as a trading and investment partner would have been seriously damaged and with it a wide range of British interests and prospects for the future [12]."

The Lib-Dem motion attacking the Government generally over its conduct of overseas aid and on Pergau, was defeated in a 305-159 division (Government majority 146 [13].)

Speaking from New York, Mr Major said last night that he had "no regrets" over the deal [14]. He was adopting a "wait and see" policy towards Malaysia's reimposition of a "buy British last" policy [15].

He refused to go into the question of retaliation, saying: "It is a matter I hope we will be able to sort out [16]."

This afternoon Mr Hurd, the Foreign Secretary, will give evidence to the Commons Foreign Affairs Select Committee, which is conducting an inquiry into the affair [17].

He will attempt to smooth ruffled Malaysian feathers while also explaining why he overruled advice from the top civil servant at the Overseas Development Administration, who described the dam as "a very bad buy" [18].

B Results of the Automatic Text Abstracting System

B.1 Abstracts for Article A (All 5 Methods)

```
Method Nr.0:

[nr. 0, rank: 0, weight: 46.5]
[nr. 2, rank: 1, weight: 39.6]
[nr. 4, rank: 2, weight: 28.2]
[nr. 5, rank: 4, weight: 25.9]
[nr. 7, rank: 3, weight: 27.9]
[nr. 10, rank: 5, weight: 23.6]
[nr. 14, rank: 6, weight: 23.4]

 COUNTRYSIDE ALLIANCE FIGHTS TRESPASS LAW.
 AN unprecedented alliance of 115 countryside, wildlife and sporting
 organisations has written to Mr Ian Lang, the Scottish Secretary,
 asking him to scrap proposals making trespass a criminal offence in
 Scotland.
 The letter says the new offence of aggravated trespass potentially
 affects every man, woman or child who wants to enjoy Scotland's
 countryside.
 It is signed by Mr Bill Murray, chairman of the Scottish Countryside
 Activities Council, Mr Fred Nelson, chairman of the Scottish Sports
 Association, and Mr Seaton Baxter, chairman of Scottish Wildlife and
 Countryside Link.
 Mr Lang has already been urged by Scotland's police constables and by
 Magnus Magnusson, chairman of Scottish Natural Heritage, the country
 recreation watchdog, to scrap the proposed new offence in Scotland.
 They say the public right to cross land in Scotland is often wrongly
 assumed to be a general right.
 The introduction of a crime of aggravated trespass would, they argue,
 lead to a confrontational situation with landowners.

Method Nr.1:

[nr. 0, rank: 4, weight: 233]
[nr. 2, rank: 0, weight: 792]
[nr. 3, rank: 1, weight: 532]
[nr. 4, rank: 2, weight: 452]
[nr. 16, rank: 6, weight: 220]
[nr. 17, rank: 5, weight: 221]
[nr. 18, rank: 3, weight: 267]

 COUNTRYSIDE ALLIANCE FIGHTS TRESPASS LAW.
 AN unprecedented alliance of 115 countryside, wildlife and sporting
 organisations has written to Mr Ian Lang, the Scottish Secretary,
 asking him to scrap proposals making trespass a criminal offence in
 Scotland.
 They express concern that parts of the Criminal Justice and Public
 Order Bill, which gives the police new powers to control the
 disturbance of hunting and grouse shooting, could be used by
 landowners to intimidate people using the outdoors for recreation.
 The letter says the new offence of aggravated trespass potentially
 affects every man, woman or child who wants to enjoy Scotland's
```

countryside.
They say the need for measures to control the disruption of hunting and grouse shooting in Scotland has yet to be demonstrated.
Even if disruption were to take place it would be relatively easy for the Government to bring in an amendment to a future Criminal Justice Bill.
They add that the present Bill makes a nonsense of a review of access to the outdoors being conducted by Scottish Natural Heritage.

Method Nr.2:

[nr. 0, rank: 1, weight: 29.1]
[nr. 2, rank: 0, weight: 39.6]
[nr. 4, rank: 2, weight: 28.2]
[nr. 5, rank: 4, weight: 25.9]
[nr. 7, rank: 3, weight: 27.9]
[nr. 10, rank: 5, weight: 23.6]
[nr. 14, rank: 6, weight: 23.4]

COUNTRYSIDE ALLIANCE FIGHTS TRESPASS LAW.
AN unprecedented alliance of 115 countryside, wildlife and sporting organisations has written to Mr Ian Lang, the Scottish Secretary, asking him to scrap proposals making trespass a criminal offence in Scotland.
The letter says the new offence of aggravated trespass potentially affects every man, woman or child who wants to enjoy Scotland's countryside.
It is signed by Mr Bill Murray, chairman of the Scottish Countryside Activities Council, Mr Fred Nelson, chairman of the Scottish Sports Association, and Mr Seaton Baxter, chairman of Scottish Wildlife and Countryside Link.
Mr Lang has already been urged by Scotland's police constables and by Magnus Magnusson, chairman of Scottish Natural Heritage, the country recreation watchdog, to scrap the proposed new offence in Scotland.
They say the public right to cross land in Scotland is often wrongly assumed to be a general right.
The introduction of a crime of aggravated trespass would, they argue, lead to a confrontational situation with landowners.

Method Nr.3:

[nr. 0, rank: 12, weight: 233]
[nr. 2, rank: 0, weight: 792]
[nr. 3, rank: 3, weight: 532]
[nr. 4, rank: 5, weight: 452]
[nr. 5, rank: 2, weight: 543]
[nr. 7, rank: 1, weight: 559]
[nr. 8, rank: 4, weight: 482]
[nr. 13, rank: 6, weight: 412]

COUNTRYSIDE ALLIANCE FIGHTS TRESPASS LAW.
AN unprecedented alliance of 115 countryside, wildlife and sporting organisations has written to Mr Ian Lang, the Scottish Secretary, asking him to scrap proposals making trespass a criminal offence in Scotland.
They express concern that parts of the Criminal Justice and Public

Order Bill, which gives the police new powers to control the disturbance of hunting and grouse shooting, could be used by landowners to intimidate people using the outdoors for recreation.
The letter says the new offence of aggravated trespass potentially affects every man, woman or child who wants to enjoy Scotland's countryside.
It is signed by Mr Bill Murray, chairman of the Scottish Countryside Activities Council, Mr Fred Nelson, chairman of the Scottish Sports Association, and Mr Seaton Baxter, chairman of Scottish Wildlife and Countryside Link.
Mr Lang has already been urged by Scotland's police constables and by Magnus Magnusson, chairman of Scottish Natural Heritage, the country recreation watchdog, to scrap the proposed new offence in Scotland.
The new alliance of outdoor groups, which includes ramblers, mountaineers, cyclists, skiers and wildlife enthusiasts, says the Government has taken no account of the different tradition and legal basis for public access in Scotland and has made no attempt to consult with any of their organisations.
The number of footpaths, tracks and wild areas where the public has secured a legal right to roam are minute when compared to the area of land to which the public presently enjoys access in Scotland - or with the official rights of way network in England and Wales.

Method Nr.4:

[nr. 0, rank: 18, weight: 0]
[nr. 2, rank: 0, weight: 39.6]
[nr. 4, rank: 1, weight: 28.2]
[nr. 5, rank: 3, weight: 25.9]
[nr. 7, rank: 2, weight: 27.9]
[nr. 9, rank: 6, weight: 22]
[nr. 10, rank: 4, weight: 23.6]
[nr. 14, rank: 5, weight: 23.4]
[nr. 16, rank: 7, weight: 22]

COUNTRYSIDE ALLIANCE FIGHTS TRESPASS LAW.
AN unprecedented alliance of 115 countryside, wildlife and sporting organisations has written to Mr Ian Lang, the Scottish Secretary, asking him to scrap proposals making trespass a criminal offence in Scotland.
The letter says the new offence of aggravated trespass potentially affects every man, woman or child who wants to enjoy Scotland's countryside.
It is signed by Mr Bill Murray, chairman of the Scottish Countryside Activities Council, Mr Fred Nelson, chairman of the Scottish Sports Association, and Mr Seaton Baxter, chairman of Scottish Wildlife and Countryside Link.
Mr Lang has already been urged by Scotland's police constables and by Magnus Magnusson, chairman of Scottish Natural Heritage, the country recreation watchdog, to scrap the proposed new offence in Scotland.
There appears to be no recognition of the fact that virtually everyone going for a walk or run in Scotland's countryside is, technically speaking, exposed to accusations of being a trespasser, say the chairmen.
They say the public right to cross land in Scotland is often wrongly assumed to be a general right.

Column	Type	Description
1	string (max. 6 chars, upper-case)	content word
2	float	word weight (tf*idf)
3	integer	word occurences in article
4	integer	global corpus frequency
5	integer	word-in-document frequency
6	string	first occuring full keyword

Table 16: Explanation of the "highest ranked word"–table.

```
The introduction of a crime of aggravated trespass would, they argue,
lead to a confrontational situation with landowners.
They say the need for measures to control the disruption of hunting
and grouse shooting in Scotland has yet to be demonstrated.
```

B.2 Abstracts for All Articles (Method 3)

Note 1: The words and numbers in the "highest ranked word"-table are explained in table 16.

Note 2: The global corpus frequencies were *not* used for the computation of the tf*idf weights; they are listed just for comparison to the word-in-document frequencies.

```
Article A:
**********

15 highest ranked words:
TRESPA :    102 :   6 :      85 :      53 : TRESPASS
SCOTLA :   72.3 :   9 :    1931 :    1365 : Scotland
COUNTR :     66 :   7 :   12731 :    7846 : COUNTRYSIDE
SCOTTI :   49.7 :   6 :    1852 :    1066 : Scottish
CRIMIN :   40.8 :   5 :    2034 :    1210 : criminal
ALLIAN :   38.4 :   3 :    1279 :     824 : ALLIANCE
PUBLIC :   37.7 :   6 :   13696 :    7785 : Public
  BILL :   36.5 :   5 :    4485 :    2848 : Bill
OFFENC :     35 :   4 :    1074 :     664 : offence
 RIGHT :   34.4 :   5 :    6026 :    4353 : right
OUTDOO :   32.1 :   3 :     121 :      96 : outdoors
CHAIRM :   31.6 :   5 :   10187 :    7489 : chairman
AGGRAV :   31.4 :   3 :     128 :     120 : aggravated
WILDLI :   30.1 :   3 :     320 :     187 : wildlife
ACCESS :   23.9 :   3 :    1881 :    1465 : access

[nr. 0, rank: 12, weight: 233]
[nr. 2, rank: 0, weight: 792]
[nr. 3, rank: 3, weight: 532]
[nr. 4, rank: 5, weight: 452]
[nr. 5, rank: 2, weight: 543]
[nr. 7, rank: 1, weight: 559]
[nr. 8, rank: 4, weight: 482]
[nr. 13, rank: 6, weight: 412]
```

COUNTRYSIDE ALLIANCE FIGHTS TRESPASS LAW.
AN unprecedented alliance of 115 countryside, wildlife and sporting organisations has written to Mr Ian Lang, the Scottish Secretary, asking him to scrap proposals making trespass a criminal offence in Scotland.
They express concern that parts of the Criminal Justice and Public Order Bill, which gives the police new powers to control the disturbance of hunting and grouse shooting, could be used by landowners to intimidate people using the outdoors for recreation.
The letter says the new offence of aggravated trespass potentially affects every man, woman or child who wants to enjoy Scotland's countryside.
It is signed by Mr Bill Murray, chairman of the Scottish Countryside Activities Council, Mr Fred Nelson, chairman of the Scottish Sports Association, and Mr Seaton Baxter, chairman of Scottish Wildlife and Countryside Link.
Mr Lang has already been urged by Scotland's police constables and by Magnus Magnusson, chairman of Scottish Natural Heritage, the country recreation watchdog, to scrap the proposed new offence in Scotland.
The new alliance of outdoor groups, which includes ramblers, mountaineers, cyclists, skiers and wildlife enthusiasts, says the Government has taken no account of the different tradition and legal basis for public access in Scotland and has made no attempt to consult with any of their organisations.
The number of footpaths, tracks and wild areas where the public has secured a legal right to roam are minute when compared to the area of land to which the public presently enjoys access in Scotland - or with the official rights of way network in England and Wales.

Article B:
****

15 highest ranked words:

HESELT	172	13	1464	625	HESELTINE
MOSES	109	9	55	24	Moses
CERTIF	73.7	8	882	417	certificate
NICHOL	70	8	913	667	Nicholas
INQUIR	62.6	8	2859	1670	inquiry
SIR	55.8	8	8117	3906	Sir
JUDGE	41.2	5	1919	1105	judge
ATTORN	37.9	4	494	320	Attorney
MINIST	37.4	6	17557	8294	ministerial
TRIAL	32.7	4	2126	1189	trial
SIGN	32.1	4	1596	1377	sign
LYELL	32	2	106	97	LYELL
AMENDE	31.7	3	120	107	amended
MATRIX	30.3	3	445	175	Matrix
ABOUT	27.7	5	28293	16394	about

[nr. 0, rank: 18, weight: 234]
[nr. 2, rank: 7, weight: 543]
[nr. 3, rank: 0, weight: 1.07e+03]
[nr. 8, rank: 4, weight: 574]
[nr. 11, rank: 1, weight: 946]
[nr. 16, rank: 6, weight: 554]
[nr. 17, rank: 5, weight: 571]

[nr. 18, rank: 2, weight: 722]
[nr. 20, rank: 3, weight: 677]

LYELL DID NOT REPORT HESELTINE CONCERN'.
THE ROLE of Sir Nicholas Lyell, the Attorney General, in the arms-to-Iraq affair was called into question again yesterday, shortly before he appears at the Scott inquiry.
Mr Alan Moses, QC, leading prosecution counsel in the Matrix Churchill trial, told the inquiry that Sir Nicholas did not inform him that Mr Michael Heseltine, President of the Board of Trade, had grave reservations about signing a public interest immunity certificate.
On his second day before the inquiry, Mr Moses said he would have wanted to know about Mr Heseltine's views, but because he had not been told of them he felt he had not represented the minister's position properly at the trial.
Mr Heseltine agreed to sign a much amended certificate after Sir Nicholas assured him that the unusual wording and limited scope of the order would be drawn to the attention of the judge by counsel, yet three days later, on Sept 10 1992, Sir Nicholas met Mr Moses but did not mention the issue.
Mr Moses's evidence to the inquiry raised doubts about a statement issued by Sir Nicholas after Mr Heseltine's appearance at the inquiry.
Sir Nicholas had said that Mr Heseltine's certificate had been expressly amended after consultation with the Attorney General and leading counsel for the prosecution.
Mr Moses told the inquiry that he had received a copy of Mr Heseltine's amended certificate but he was not told what changes Mr Heseltine had made, or their significance.
He added: I can well see a minister would be very ill-advised to sign a certificate unless he was quite satisfied as to the reasoning of the prosecution for fear, as I think Mr Heseltine feared, that he would be accused of withholding documents which a judge has ordered to be disclosed.

Article C:
****

15 highest ranked words:

EC	110	10	7251	2670	EC
MAJOR	49.5	5	10273	5683	MAJOR
ENLARG	47.2	5	397	334	enlarge
EUROPE	44.2	7	17043	7650	Europe
BRITAI	37.1	6	15506	8713	Britain's
VOTING	35.4	4	819	608	voting
VOTES	34.2	4	1184	815	votes
COMPRO	34.2	4	1003	820	compromise
MINIST	31.1	5	17557	8294	ministers
EURO-S	28.3	3	457	335	Euro-sceptics
WHEN	28.1	5	25193	15082	when
ISSUE	27.8	4	5476	3989	issue
WITHIN	27.5	4	5249	4360	within
HURD	26.5	3	1214	609	Hurd
BRUSSE	23.9	3	2622	1442	Brussels

[nr. 0, rank: 15, weight: 216]

```
[nr. 2, rank: 0, weight: 752]
[nr. 6, rank: 1, weight: 457]
[nr. 8, rank: 4, weight: 383]
[nr. 14, rank: 5, weight: 358]
[nr. 15, rank: 2, weight: 446]
[nr. 18, rank: 3, weight: 411]
[nr. 22, rank: 6, weight: 334]
```

MAJOR IS READY FOR LONG FIGHT WITH EC.
Britain's rift with the rest of Europe deepened last night after EC foreign ministers failed to break the deadlock over voting rights and Mr John Major, adopting a strident anti-Brussels stance, signalled that the Government was digging in for a long battle.
Yesterday's talks in Brussels broke up when it became clear there was no compromise acceptable to Britain and Spain - the only other country opposing an increase in the number of votes needed to block EC legislation.
Britain's European partners said that failure to solve the voting issue within the next few days could derail plans to enlarge the Community from 12 members to 16 with the inclusion of Austria, Sweden, Norway and Finland next year.
Delighted Tory Euro-sceptics claimed that Mr Major, faced with the prospect of a Conservative civil war or a clash with Europe, had chosen to battle with the EC.
But there was dismay among the party's European wing, who fear enlargement of the EC is threatened and that Mr Major is jeopardising compromise within the party over the issue.
Mr Kenneth Clarke, the Chancellor, said during a visit to Sweden that London felt the voting issue should be uncoupled from the enlargement process and deferred until 1996, when EC leaders are to hold an inter-governmental conference.
Britain is demanding a protocol to the enlargement treaty which would commit the EC not to over-ride a vote of 23 within a still unspecified time limit.

Article D:
****

15 highest ranked words:

WHITEW	66.6	4	164	64	WHITEWATER
CLINTO	64	8	4806	1404	Clinton's
PUBLIC	47.2	5	13696	7785	PUBLIC
SENATE	44.4	3	323	217	SENATE
CONGRE	40.9	5	1833	1182	Congressional
INQUIR	35.2	3	2859	1670	INQUIRY
HEARIN	33.5	4	1318	959	hearings
PRESID	33.2	5	10165	5503	President
WASHIN	31.1	4	2962	1782	Washington
SENATO	29.5	3	353	226	Senator
DEMOCR	29.5	4	4843	2664	Democratic
DISCLO	24.2	3	1640	1302	disclosed
REPUBL	24.1	3	2307	1364	Republican
STAGE	23.5	2	1912	1677	STAGE
ARKANS	20.9	2	212	122	Arkansas

```
[nr. 0, rank: 10, weight: 226]
[nr. 2, rank: 0, weight: 729]
[nr. 4, rank: 3, weight: 338]
[nr. 8, rank: 5, weight: 311]
[nr. 10, rank: 1, weight: 378]
[nr. 15, rank: 2, weight: 354]
[nr. 17, rank: 4, weight: 313]
```

```
 INTERNATIONAL - SENATE TO STAGE PUBLIC INQUIRY ON WHITEWATER.
 WASHINGTON began preparing yesterday for one of the biggest
 Congressional spectaculars since the Iran-Contra inquiry following a
 Senate decision to hold public hearings on President Clinton's
 Whitewater property dealings.
 But first they must agree a timetable with the Democrats who, despite
 having caved in, remain deeply uneasy at the prospect of seeing the
 President's pre - election record of private financial manoeuvres
 subjected to public scrutiny.
 Another White House insider, Presidential Counsellor David Gergen,
 expressed dismay at the overheated atmosphere in Washington and the
 vehemence of Mr Clinton's critics, which he attributed largely to his
 ambitious reform programme.
 Even as the Clinton team rallied to the banner, however, new
 information was leaking out about the Clintons' family finances which,
 while not suggesting any improprieties, does indicate how the couple
 benefited from the cronyism in their native state of Arkansas.
 Yesterday, the leader of the Senate's Democratic majority, Senator
 George Mitchell, and his Republican opposite number, Senator Robert
 Dole, were discussing with their inner circles what an appropriate
 timetable for the hearings will be.
 But with the collapse of unanimity within the party as more members
 concluded that full public disclosure of the Whitewater facts was
 preferable to perceived secretiveness, the leadership was forced to
 give way to Republican demands.
```

```
Article E:
**********

15 highest ranked words:
TROOPS :   83.8 :   7 :    3108 :     1436 : TROOPS
BOSNIA :   72.1 :   6 :    6794 :     1400 : BOSNIA
CEASEF :   64.7 :   7 :     672 :      409 : ceasefire
    UN :   47.6 :   6 :    5469 :     1517 : UN
  HELP :     34 :   5 :    6059 :     4685 : help
DEFENC :   33.2 :   3 :    4638 :     2617 : DEFENCE
COMMAN :   31.9 :   4 :    2317 :     1434 : Commander
ANNOUN :   29.4 :   3 :    7704 :     6066 : ANNOUNCING
GOVERN :   25.9 :   3 :   31775 :    13281 : GOVERNMENT
 SPLIT :   25.2 :   3 :    1139 :      952 : Split
SOLDIE :     25 :   3 :    1772 :     1025 : soldiers
  HOUR :   24.6 :   2 :    1437 :     1166 : HOUR
 FRONT :   23.7 :   3 :    1844 :     1538 : front
COLDST :   23.5 :   2 :      42 :       33 : Coldstream
CROAT- :   23.2 :   2 :      45 :       38 : Croat-Muslim

[nr. 0, rank: 6, weight: 349]
[nr. 3, rank: 2, weight: 488]
```

```
[nr. 5, rank: 1, weight: 549]
[nr. 8, rank: 5, weight: 386]
[nr. 9, rank: 3, weight: 457]
[nr. 13, rank: 4, weight: 410]
[nr. 14, rank: 0, weight: 581]
```

INTERNATIONAL - TROOPS LEAVE FOR BOSNIA WITHIN HOUR OF DEFENCE SECRETARY ANNOUNCING GOVERNMENT DECISION.
THE British troops who began arriving in Split last night will help the United Nations monitor a delicate Croat-Muslim ceasefire in Bosnia.
The first VC-10 load of troops, an advance party from the 1st Bn The Duke of Wellington's Regiment, left RAF Brize Norton, Oxon, within an hour of Mr Malcolm Rifkind, the Defence Secretary, announcing to the Commons that the Government had decided to send the extra forces to support the UN's request for 10,650.
By then, 792 soldiers, including engineers and logistics troops, along with 113 men with 15 Scimitar tracked reconnaissance vehicles from D Squadron of The Light Dragoons at Hohne, Germany, will have been deployed to Bosnia to join 2,300 already there.
Explaining the Government's change of mind about extra troops, Mr Rifkind told the Commons that the ceasefires had brought new opportunities and responsibilities that were greatly stretching the ability of the Coldstream Guards battalion group to fulfil its UN mission.
The Duke of Wellington's will provide Lt-Gen Sir Michael Rose, the UN Commander in Bosnia, with his most powerful new body to help cement the Croat-Muslim ceasefire.
The troops will work to the complex ceasefire map drawn up by a joint commission of Bosnian army and Croat HVO militia commanders under the chairmanship of Brig John Reith, Commander of the new UN Sector SouthWest in Bosnia.

Article F:
****

15 highest ranked words:

PERGAU :	69.1 :	4 :	119 :	42 :	PERGAU
MALAYS :	58.7 :	6 :	601 :	236 :	Malaysian
AID :	55.3 :	7 :	2983 :	1568 :	aid
SIR :	41.9 :	6 :	8117 :	3906 :	Sir
CIRCLE :	41.4 :	3 :	509 :	424 :	CIRCLE'
THATCH :	37.4 :	3 :	2090 :	1043 :	THATCHER
DAM :	32.3 :	3 :	182 :	89 :	Dam
TRADE :	27.7 :	4 :	6881 :	4086 :	trade
FOREIG :	27.6 :	4 :	6813 :	4211 :	foreign
DAVID :	27.5 :	4 :	5278 :	4385 :	David
ARMS :	25.2 :	3 :	1496 :	960 :	arms
BRITAI :	24.7 :	4 :	15506 :	8713 :	Britain
TIPPIN :	24 :	2 :	52 :	26 :	Tipping
AFFAIR :	22.7 :	3 :	2623 :	2146 :	affairs
PROJEC :	22.4 :	3 :	3759 :	2403 :	project

```
[nr. 0, rank: 15, weight: 160]
[nr. 2, rank: 0, weight: 581]
[nr. 3, rank: 1, weight: 568]
[nr. 4, rank: 2, weight: 501]
```

[nr. 7, rank: 6, weight: 342]
[nr. 9, rank: 4, weight: 412]
[nr. 10, rank: 3, weight: 412]
[nr. 12, rank: 5, weight: 367]

THATCHER CIRCLE' IN PERGAU ROW.
The close circle around Lady Thatcher was criticised yesterday by Sir David Steel, the Liberal Democrats' foreign affairs spokesman, for involvement in the Pergau Dam project.
Speaking in an acrimonious Commons debate on the #234 million of British aid for the dam, Sir David said: In the course of batting for Britain, which our former Prime Minister did extremely well, her close circle often seems to have been involved.
In this case Sir Tim Bell, well known as her PR adviser, is also adviser to the Malaysian Prime Minister and to Tam Sri Armugam, who controls GEC Malaysia, heavily involved in several of the contracts under the deal.
Opening the debate, Sir David said the only possible explanation for the funding of the project was the Government's eagerness to lubricate the trade channels with Malaysia, especially the arms trade.
Seeking to establish a link between aid and arms sales, he pointed out that Oman, whose GNP per head is higher than Portugal's had had its aid from Britain doubled since 1979 and it currently ranks third largest in the list of purchasers of arms from Britain.
Referring to an unhealthy dominance of Tory Party benefactors among firms in the Pergau project, Sir David said: Companies linked by such donations have been the main beneficiaries to the tune of 42 per cent of the Aid and Trade Provision.
Mr Alistair Goodlad, a Foreign Office minister, ridiculed Sir David's figures, adding: It is ridiculous to suggest that a country where there is a defence sales relationship should, because of that relationship, be ineligible for aid for trade provision.

C Results of the Empirical Study

C.1 Significance Tables for Sentence Score Correlations

	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
HS2		n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	n.s.	n.s.	n.s.
HS4			n.s.	n.s.	***	***	*	*	*	*	n.s.	n.s.	n.s.
HS3				n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
HS8					*	n.s.	*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
HS9						***	***	***	***	*	n.s.	***	n.s.
HS1							***	***	***	***	n.s.	***	n.s.
HS5								*	***	*	n.s.	*	n.s.
HS12									***	n.s.	n.s.	*	n.s.
HS11										n.s.	n.s.	*	n.s.
HS13											***	***	n.s.
HS10												n.s.	n.s.
HS14													n.s.

Table 17: Significance of sentence score correlation: Article A

	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
HS2		***	n.s.	***	***	n.s.	***	n.s.	***	*	n.s.	***	*
HS4			n.s.	***	***	n.s.	***	n.s.	*	n.s.	n.s.	***	***
HS3				n.s.	n.s.	n.s.	*	*	n.s.	n.s.	n.s.	n.s.	n.s.
HS8					*	n.s.	*	n.s.	***	*	n.s.	*	*
HS9						n.s.	***	n.s.	*	n.s.	n.s.	*	*
HS1							n.s.	***	n.s.	*	*	n.s.	n.s.
HS5								n.s.	n.s.	n.s.	n.s.	*	n.s.
HS12									n.s.	***	n.s.	n.s.	n.s.
HS11										*	n.s.	n.s.	*
HS13											n.s.	n.s.	n.s.
HS10												n.s.	n.s.
HS14													*

Table 18: Significance of sentence score correlation: Article B

	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
HS2		***	***	***	***	***	*	*	***	***	***	***	***
HS4			***	***	***	***	*	n.s.	***	***	***	***	***
HS3				***	n.s.	***	***	n.s.	***	***	***	***	***
HS8					n.s.	*	*	n.s.	***	***	***	*	***
HS9						***	n.s.	n.s.	***	*	***	***	*
HS1							***	***	***	***	***	***	***
HS5								***	n.s.	*	n.s.	*	*
HS12									n.s.	n.s.	n.s.	n.s.	*
HS11										***	***	*	***
HS13											***	***	***
HS10												*	***
HS14													***

Table 19: Significance of sentence score correlation: Article C

	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
HS2		n.s.	n.s.	n.s.	*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
HS4			***	***	n.s.	n.s.	n.s.	***	n.s.	n.s.	*	n.s.	n.s.
HS3				***	n.s.	*	***	***	***	n.s.	***	n.s.	*
HS8					n.s.	n.s.	n.s.	***	***	n.s.	***	n.s.	n.s.
HS9						***	n.s.	n.s.	n.s.	*	n.s.	*	n.s.
HS1							***	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
HS5								n.s.	n.s.	n.s.	*	n.s.	n.s.
HS12									*	n.s.	***	n.s.	n.s.
HS11										n.s.	*	n.s.	n.s.
HS13											n.s.	n.s.	n.s.
HS10												n.s.	*
HS14													n.s.

Table 20: Significance of sentence score correlation: Article D

	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
HS2		n.s.	*	n.s.	n.s.	***	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.
HS4			n.s.	***	*	*	***	n.s.	*	n.s.	n.s.	***	***
HS3				n.s.	*	n.s.	*	*	n.s.	n.s.	n.s.	***	n.s.
HS8					n.s.	*	n.s.	n.s.	***	n.s.	n.s.	***	*
HS9						*	***	n.s.	n.s.	n.s.	n.s.	n.s.	*
HS1							n.s.	n.s.	***	n.s.	n.s.	n.s.	***
HS5								n.s.	n.s.	n.s.	n.s.	*	n.s.
HS12									n.s.	***	n.s.	n.s.	n.s.
HS11										n.s.	n.s.	*	***
HS13											n.s.	n.s.	n.s.
HS10												n.s.	n.s.
HS14													n.s.

Table 21: Significance of sentence score correlation: Article E

	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
HS2		***	*	n.s.	***	*	*	*	n.s.	***	*	n.s.	n.s.
HS4			n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	***	n.s.	n.s.	n.s.
HS3				n.s.	*	***	*	*	n.s.	*	*	*	***
HS8					n.s.	n.s.	n.s.	*	***	n.s.	n.s.	***	n.s.
HS9						***	*	n.s.	n.s.	***	***	n.s.	n.s.
HS1							n.s.	n.s.	n.s.	*	n.s.	n.s.	n.s.
HS5								***	n.s.	***	*	n.s.	n.s.
HS12									n.s.	***	n.s.	n.s.	*
HS11										n.s.	n.s.	***	n.s.
HS13											*	n.s.	*
HS10												n.s.	n.s.
HS14													*

Table 22: Significance of sentence score correlation: Article F

C.2 Significance Tables for Chi Square Evaluations

	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
HS2		***	n.s.	n.s.	n.s.	*	n.s.	***	*	***	n.s.	*	n.s.
HS4			***	***	***	***	***	n.s.	***	***	n.s.	***	***
HS3				*	***	n.s.	***	*	n.s.	n.s.	*	***	***
HS8					n.s.	*	***	***	*	*	n.s.	***	n.s.
HS9						***	***	n.s.	n.s.	n.s.	n.s.	***	n.s.
HS1							*	*	n.s.	***	***	***	*
HS5								n.s.	n.s.	n.s.	n.s.	*	n.s.
HS12									n.s.	***	n.s.	n.s.	n.s.
HS11										n.s.	***	n.s.	***
HS13											***	*	n.s.
HS10												n.s.	*
HS14													n.s.

Table 23: Significance of sentence score chi squares (5/4 vs. 3/2/1): All 6 articles

	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
HS2		***	*	*	***	***	*	***	*	***	*	***	n.s.
HS4			***	***	*	***	***	n.s.	***	***	*	n.s.	***
HS3				***	***	***	***	***	***	n.s.	***	*	***
HS8					***	*	*	***	***	n.s.	***	*	n.s.
HS9						***	***	n.s.	*	***	n.s.	***	n.s.
HS1							***	n.s.	***	*	*	n.s.	***
HS5								***	*	***	*	n.s.	***
HS12									***	***	*	n.s.	n.s.
HS11										*	*	*	***
HS13											***	n.s.	n.s.
HS10												n.s.	***
HS14													*

Table 24: Significance of sentence selections' chi squares: All 6 articles

C.3 Optimal Abstracts According to Top Ranked Mean Scores

Article A: Countryside Alliance Fights Trespass Law.

AN unprecedented alliance of 115 countryside, wildlife and sporting organisations has written to Mr Ian Lang, the Scottish Secretary, asking him to scrap proposals making trespass a criminal offence in Scotland [1].

They express concern that parts of the Criminal Justice and Public Order Bill, which gives the police new powers to control the disturbance of hunting and grouse shooting, could be used by landowners to "intimidate" people using the outdoors for recreation [2].

The new alliance of outdoor groups, which includes ramblers, mountaineers, cyclists, skiers and wildlife enthusiasts, says the Government has taken no account of the different tradition and legal basis for public access in Scotland and has made no attempt to consult with any of their organisations [7]. "There appears to be no recognition of the fact that virtually everyone going for a walk or run in Scotland's countryside is, technically speaking, exposed to accusations of being a trespasser," say the chairmen [8].

They say the public right to cross land in Scotland is often wrongly assumed to be a general right [9].

The number of footpaths, tracks and wild areas where the public has secured a legal right to roam are "minute" when compared to the area of land to which the public presently enjoys access in Scotland - or with the official rights of way network in England and Wales [12].

Article B: Lyell 'Did Not Report Heseltine Concern'.

THE ROLE of Sir Nicholas Lyell, the Attorney General, in the arms-to-Iraq affair was called into question again yesterday, shortly before he appears at the Scott inquiry [1].

Mr Alan Moses, QC, leading prosecution counsel in the Matrix Churchill trial, told the inquiry that Sir Nicholas did not inform him that Mr Michael Heseltine, President of the Board of Trade, had grave reservations about signing a public interest immunity certificate [2].

The trial of three former executives of Matrix Churchill, charged with illegally exporting arms-making machinery to Iraq, collapsed in November 1992, after certificates were overturned by the judge, releasing to the defence Government papers showing ministerial knowledge of the trade [3].

Last month the inquiry was told that Mr Heseltine had refused to sign a gagging order to withhold Whitehall papers from the defence in the case because he feared he would be taking part in a "cover-up" [8].

Sir Nicholas told Mr Heseltine that despite his concerns it was his ministerial "duty" to sign [9]. Mr Heseltine agreed to sign a much amended certificate after Sir Nicholas assured him that the "unusual wording" and "limited scope" of the order would be drawn to the attention of the judge by counsel, yet three days later, on Sept 10 1992, Sir Nicholas met Mr Moses but did not mention the issue [10].

Article C: Major Is Ready For Long Fight With EC.

Britain's rift with the rest of Europe deepened last night after EC foreign ministers failed to break the deadlock over voting rights and Mr John Major, adopting a strident anti-Brussels stance, signalled that the Government was digging in for a long battle [1].

Yesterday's talks in Brussels broke up when it became clear there was no compromise acceptable to Britain and Spain - the only other country opposing an increase in the number of votes needed to block EC legislation [5].

Britain's European partners said that failure to solve the voting issue within the next few days could derail plans to enlarge the Community from 12 members to 16 with the inclusion of Austria, Sweden, Norway and Finland next year [7].

The voting row centres on whether the number of votes in the EC's ruling Council of Ministers needed to block legislation should be increased from 23 to 27 when the four new countries join [15].

There remains strong opposition within the Cabinet to giving way on 27 [16].

Emerging from the Brussels talks, Mr Hurd said he had rejected a compromise proposal, which would allow for a delay of one or two months in cases where there were 23 votes against a proposal, but not the full 27 [18].

Britain is demanding a protocol to the enlargement treaty which would commit the EC not to over-ride a vote of 23 within a still unspecified time limit [21].

Article D: International - Senate To Stage Public Inquiry On Whitewater.

WASHINGTON began preparing yesterday for one of the biggest Congressional spectaculars since the Iran-Contra inquiry following a Senate decision to hold public hearings on President Clinton's Whitewater property dealings [1].

But first they must agree a timetable with the Democrats who, despite having caved in, remain deeply uneasy at the prospect of seeing the President's pre- election record of private financial manoeuvres subjected to public scrutiny [3].

The risk, he said, is that the Whitewater controversy will paralyse the Presidency [8].

Even as the Clinton team rallied to the banner, however, new information was leaking out about the Clintons' family finances which, while not suggesting any improprieties, does indicate how the couple benefited from the "cronyism" in their native state of Arkansas [9].

Mrs Hillary Clinton, the New York Times disclosed, made nearly #660,000 by trading in cattle futures in the late 1970s on the advice of a lawyer friend who worked for Tyson Foods, one of Arkansas's biggest companies [10].

But with the collapse of unanimity within the party as more members concluded that full public disclosure of the Whitewater facts was preferable to perceived secretiveness, the leadership was forced to give way to Republican demands [16].

Article E: International - Troops Leave For Bosnia Within Hour Of Defence Secretary Announcing Government Decision.

THE British troops who began arriving in Split last night will help the United Nations monitor a delicate Croat-Muslim ceasefire in Bosnia [1].

Explaining the Government's change of mind about extra troops, Mr Rifkind told the Commons that the ceasefires had brought new opportunities and responsibilities that were greatly stretching the ability of the Coldstream Guards battalion group to fulfil its UN mission [7].

Although the UK contribution to the region is already a large one, a further UK contribution at this stage as part of a co-ordinated international effort would help to make the difference between success and failure for the ceasefires [10]."

The troops will work to the complex ceasefire map drawn up by a joint commission of Bosnian army and Croat HVO militia commanders under the chairmanship of Brig John Reith, Commander of the new UN Sector SouthWest in Bosnia [12].

British officials say the map shows a ceasefire to be policed along 125 miles of front lines between the two sides, some of which cross high mountain ridges [13].

The soldiers are needed to patrol no man's land, help take weapons to collection sites and man checkpoints and observation posts [14].

Next week, the UN soldiers will begin overseeing the disengagement of forces [15].

Article F: 'Thatcher Circle' In Pergau Row.

The "close circle" around Lady Thatcher was criticised yesterday by Sir David Steel, the Liberal Democrats' foreign affairs spokesman, for involvement in the Pergau Dam project [1].

Speaking in an acrimonious Commons debate on the #234 million of British aid for the dam, Sir David said: "In the course of batting for Britain, which our former Prime Minister did extremely well, her close circle often seems to have been involved [2].

"In this case Sir Tim Bell, well known as her PR adviser, is also adviser to the Malaysian Prime Minister and to Tam Sri Armugam, who controls GEC Malaysia, heavily involved in several of the contracts under the deal [3].

Opening the debate, Sir David said the only possible explanation for the funding of the project was the Government's "eagerness to lubricate the trade channels with Malaysia, especially the arms trade" [6].

Referring to an "unhealthy dominance" of Tory Party benefactors among firms in the Pergau project, Sir David said: "Companies linked by such donations have been the main beneficiaries to the tune of 42 per cent of the Aid and Trade Provision [9]."

The Lib-Dem motion attacking the Government generally over its conduct of overseas aid and on Pergau, was defeated in a 305-159 division (Government majority 146 [13].)

D Tables Concerning the Man–Machine Comparison

	M0	M1	M2	M3	M4	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
M0		0.52	0.89	0.68	0.9	0.38	0.038	0.44	0.087	0	0.32	0.15	0.48	0.11	0.2	0.43	0.11	-0.31
M1			0.6	0.54	0.52	0.3	0.2	0.61	0.14	0.15	0.42	0.29	0.43	0.32	0.71	0.86	0.27	0.13
M2				0.72	0.89	0.39	0.048	0.41	0.0061	0	0.32	0.073	0.53	0.17	0.23	0.53	0.16	-0.21
M3					0.84	0.58	0.22	0.35	-0.11	0.057	0.39	0.077	0.66	0.19	0.22	0.39	0.054	-0.19
M4						0.35	0.086	0.31	0.018	-0.076	0.31	-0.0097	0.52	0.087	0.14	0.36	0.028	-0.28
HS2	0.5	0.5	0.5	0.57	0.38		0.027	0.12	-0.071	0.22	0.31	-0.026	0.43	0.04	0.24	0.33	0.077	-0.15
HS4	0.17	0.33	0.17	0.57	0.25			0.058	0.13	0.6	0.67	0.47	0.5	0.43	0.55	0.15	0.34	0.35
HS3	0.5	0.5	0.5	0.43	0.38				0.19	0.071	0.14	0.34	0.28	0.12	0.39	0.33	0.21	-0.2
HS8	0.33	0.33	0.33	0.29	0.38					0.54	0.17	0.53	0.05	0.39	0.19	-0.0041	0.22	-0.17
HS9	0.17	0.33	0.17	0.57	0.12						0.65	0.63	0.56	0.56	0.55	0.18	0.7	0.16
HS1	0.33	0.5	0.33	0.57	0.5							0.61	0.73	0.69	0.69	0.39	0.59	0.19
HS5	0.33	0.33	0.33	0.43	0.38								0.45	0.84	0.42	0.27	0.53	0.22
HS12	0.33	0.33	0.33	0.57	0.25									0.57	0.36	0.33	0.5	-0.067
HS11	0.33	0.5	0.33	0.43	0.5										0.4	0.35	0.54	0.11
HS13	0.17	0.67	0.17	0.43	0.25											0.71	0.59	0.3
HS10	0.5	1	0.5	0.43	0.5												0.33	0.22
HS14	0.17	0.33	0.17	0.43	0.25													0.33
HS15	0.17	0.67	0.17	0.57	0.25													
MEAN	0.31	0.49	0.31	0.48	0.34													

Table 25: Pearson's r and Precision (left below diagonal) for Article A.

	M0	M1	M2	M3	M4	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
M0		*	***	***	***	n.s.	n.s.	*	n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	*	n.s.	n.s.
M1			***	*	*	n.s.	n.s.	***	n.s.	n.s.	*	n.s.	*	n.s.	***	***	n.s.	n.s.
M2				***	***	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	*	n.s.	n.s.
M3					***	***	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	***	n.s.	n.s.	n.s.	n.s.	n.s.
M4						n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	n.s.	n.s.	n.s.
HS2	0.43	0.43	0.43	0.57	0.43		n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	n.s.	n.s.	n.s.
HS4	0.2	0.4	0.2	0.8	0.4			n.s.	n.s.	***	***	*	*	*	*	n.s.	n.s.	n.s.
HS3	0.6	0.6	0.6	0.6	0.6				n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
HS8	0.4	0.4	0.4	0.4	0.6					*	n.s.	*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
HS9	0.17	0.33	0.17	0.67	0.17						***	***	***	***	*	n.s.	***	n.s.
HS1	0.29	0.43	0.29	0.57	0.57							***	***	***	***	n.s.	***	n.s.
HS5	0.4	0.4	0.4	0.6	0.6								*	***	*	n.s.	*	n.s.
HS12	0.5	0.5	0.5	1	0.5									***	n.s.	n.s.	*	n.s.
HS11	0.33	0.5	0.33	0.5	0.67										n.s.	n.s.	*	n.s.
HS13	0.2	0.8	0.2	0.6	0.4											***	***	n.s.
HS10	0.38	0.75	0.38	0.38	0.5												n.s.	n.s.
HS14	0.2	0.4	0.2	0.6	0.4													n.s.
HS15	0.14	0.57	0.14	0.57	0.29													
MEAN	0.33	0.5	0.33	0.6	0.47													

Table 26: Significance of Pearson's r and Recall (left below diagonal) for Article A.

	M0	M1	M2	M3	M4	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
M0		0.27	0.88	0.36	0.74	0.53	0.23	-0.078	0.13	0.28	-0.047	0.24	0.021	0.18	0.049	-0.14	0.38	0.38
M1			0.21	0.57	0.29	0.32	0.2	0.53	0.35	0.18	0.31	0.25	0.42	0.062	0.22	0.76	0.45	0.24
M2				0.38	0.78	0.5	0.052	0.069	0.17	0.21	-0.047	0.21	0.19	0.081	0.15	-0.24	0.25	0.21
M3					0.46	0.53	0.52	0.59	0.49	0.31	0.31	0.51	0.38	0.31	0.12	0.19	0.41	0.54
M4						0.36	0.087	0.26	0.2	0.15	0.23	0.32	0.26	-0.0093	0.11	-0.19	0.29	0.25
HS2	0.6	0.5	0.57	0.38	0.57		0.63	0.05	0.64	0.68	0.29	0.57	0.2	0.62	0.46	0.095	0.63	0.49
HS4	0.4	0.17	0.43	0.25	0.43			0.17	0.6	0.66	0.18	0.68	-0.012	0.5	0.14	0.11	0.57	0.54
HS3	0.2	0.67	0.29	0.5	0.29				0.23	0.15	0.31	0.51	0.5	-0.26	-0.013	0.34	0.29	0.011
HS8	0.4	0.67	0.43	0.38	0.43					0.48	0.33	0.4	0.3	0.66	0.44	0.26	0.5	0.47
HS9	0.4	0.5	0.43	0.38	0.43						0.33	0.61	-0.01	0.45	0.28	0.2	0.51	0.48
HS1	0.2	0.5	0.14	0.25	0.14							0.16	0.58	0.31	0.49	0.45	0.19	0.052
HS5	0.4	0.33	0.43	0.5	0.43								0.3	0.19	0.33	0.044	0.44	0.33
HS12	0.4	0.5	0.43	0.5	0.43									0.14	0.58	0.34	-0.17	-0.19
HS11	0.4	0.5	0.57	0.5	0.57										0.52	0.1	0.27	0.45
HS13	0.2	0.5	0.29	0.38	0.29											0.29	0.038	0.11
HS10	0.4	1	0.29	0.5	0.29												0.32	0.016
HS14	0.8	0.5	0.71	0.62	0.71													0.43
HS15	0.2	0.33	0.29	0.38	0.29													
MEAN	0.38	0.51	0.41	0.42	0.41													

Table 27: Pearson's r and Precision (left below diagonal) for Article B.

	M0	M1	M2	M3	M4	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
M0		n.s.	***	n.s.	***	*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
M1			n.s.	***	n.s.	n.s.	n.s.	***	n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	***	*	n.s.
M2				n.s.	***	*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
M3					*	***	*	***	*	n.s.	n.s.	*	n.s.	n.s.	n.s.	n.s.	*	***
M4						n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
HS2	0.5	0.5	0.67	0.5	0.67		***	n.s.	***	***	n.s.	***	n.s.	***	*	n.s.	***	*
HS4	0.33	0.17	0.5	0.33	0.5			n.s.	***	***	n.s.	***	n.s.	*	n.s.	n.s.	***	***
HS3	0.17	0.67	0.33	0.67	0.33				n.s.	n.s.	n.s.	*	*	n.s.	n.s.	n.s.	n.s.	n.s.
HS8	0.29	0.57	0.43	0.43	0.43					*	n.s.	*	n.s.	***	*	n.s.	*	*
HS9	0.29	0.43	0.43	0.43	0.43						n.s.	***	n.s.	*	n.s.	n.s.	*	*
HS1	0.17	0.5	0.17	0.33	0.17							n.s.	***	n.s.	*	*	n.s.	n.s.
HS5	0.4	0.4	0.6	0.8	0.6								n.s.	n.s.	n.s.	n.s.	*	n.s.
HS12	0.29	0.43	0.43	0.57	0.43									n.s.	***	n.s.	n.s.	n.s.
HS11	0.29	0.43	0.57	0.57	0.57										*	n.s.	n.s.	*
HS13	0.17	0.5	0.33	0.5	0.33											n.s.	n.s.	n.s.
HS10	0.33	1	0.33	0.67	0.33												n.s.	n.s.
HS14	0.57	0.43	0.71	0.71	0.71													*
HS15	0.2	0.4	0.4	0.6	0.4													
MEAN	0.31	0.49	0.45	0.55	0.45													

Table 28: Significance of Pearson's r and Recall (left below diagonal) for Article B.

	M0	M1	M2	M3	M4	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
M0		0.36	0.94	0.57	0.58	0.017	0.34	0.011	-0.02	0.3	0.2	-0.3	0.012	0.15	0	0.054	0.32	0.25
M1			0.45	0.59	0.45	0.45	0.31	0.41	0.37	0.19	0.35	0.29	0.29	0.29	0.46	0.39	0.25	0.29
M2				0.61	0.71	0.12	0.32	0.11	0.03	0.32	0.3	-0.27	0.078	0.18	0.15	0.083	0.42	0.3
M3					0.64	0.38	0.38	0.4	0.43	0.25	0.45	0.16	0.37	0.31	0.4	0.29	0.4	0.5
M4						0.21	0.29	0.36	0.26	0.27	0.24	-0.066	0.1	0.23	0.31	0.16	0.51	0.48
HS2	0.33	0.33	0.33	0.43	0.38		0.69	0.67	0.69	0.52	0.73	0.5	0.38	0.64	0.74	0.7	0.62	0.63
HS4	0.33	0.33	0.33	0.43	0.38			0.56	0.65	0.68	0.65	0.38	0.32	0.92	0.64	0.81	0.53	0.69
HS3	0.33	0.5	0.33	0.57	0.5				0.71	0.37	0.51	0.59	0.36	0.51	0.76	0.55	0.56	0.7
HS8	0.33	0.33	0.33	0.29	0.25					0.37	0.42	0.43	0.25	0.67	0.84	0.77	0.49	0.54
HS9	0.33	0.17	0.33	0.43	0.38						0.64	0.27	0.14	0.57	0.47	0.53	0.67	0.46
HS1	0.33	0.33	0.33	0.57	0.38							0.55	0.57	0.54	0.56	0.54	0.74	0.58
HS5	0.17	0.5	0.17	0.29	0.25								0.54	0.32	0.45	0.37	0.37	0.38
HS12	0.17	0.17	0.17	0.29	0.12									0.29	0.3	0.33	0.2	0.44
HS11	0.5	0.5	0.5	0.57	0.5										0.71	0.9	0.4	0.73
HS13	0.33	0.33	0.33	0.57	0.38											0.82	0.57	0.64
HS10	0.33	0.33	0.33	0.43	0.25												0.43	0.68
HS14	0.33	0.33	0.33	0.43	0.38													0.51
HS15	0.5	0.33	0.5	0.57	0.5													
MEAN	0.33	0.35	0.33	0.45	0.36													

Table 29: Pearson's r and Precision (left below diagonal) for Article C.

	M0	M1	M2	M3	M4	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
M0		n.s.	***	***	***	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
M1			*	***	*	*	n.s.	*	*	n.s.	n.s.	n.s.	n.s.	n.s.	*	*	n.s.	n.s.
M2				***	***	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.
M3					***	*	*	*	*	n.s.	*	n.s.	*	n.s.	*	n.s.	*	*
M4						n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	***	*
HS2	0.4	0.4	0.4	0.6	0.6		***	***	***	***	***	*	*	***	***	***	***	***
HS4	0.29	0.29	0.29	0.43	0.43			***	***	***	***	*	n.s.	***	***	***	***	***
HS3	0.29	0.43	0.29	0.57	0.57				***	n.s.	***	***	n.s.	***	***	***	***	***
HS8	0.5	0.5	0.5	0.5	0.5					n.s.	*	*	n.s.	***	***	***	*	***
HS9	0.33	0.17	0.33	0.5	0.5						***	n.s.	n.s.	***	*	***	***	*
HS1	0.29	0.29	0.29	0.57	0.43							***	***	***	***	***	***	***
HS5	0.17	0.5	0.17	0.33	0.33								***	n.s.	*	n.s.	*	*
HS12	0.33	0.33	0.33	0.67	0.33									n.s.	n.s.	n.s.	n.s.	*
HS11	0.43	0.43	0.43	0.57	0.57										***	***	*	***
HS13	0.33	0.33	0.33	0.67	0.5											***	***	***
HS10	0.29	0.29	0.29	0.43	0.29												*	***
HS14	0.4	0.4	0.4	0.6	0.6													***
HS15	0.5	0.33	0.5	0.67	0.67													
MEAN	0.35	0.36	0.35	0.55	0.49													

Table 30: Significance of Pearson's r and Recall (left below diagonal) for Article C.

	M0	M1	M2	M3	M4	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
M0		0.65	0.72	0.26	0.17	0.053	0.36	0.33	0.17	0.46	0.7	0.41	0.37	-0.023	0.54	0.3	0.3	-0.23
M1			0.68	0.7	0.53	0.42	0.55	0.28	0.11	0.38	0.47	0.19	0.57	0.094	0.61	0.41	0.35	-0.015
M2				0.6	0.61	0.077	0.47	0.43	0.33	0.38	0.47	0.45	0.51	0.21	0.48	0.37	0.56	-0.1
M3					0.79	0.37	0.68	0.36	0.41	0.15	0.28	0.31	0.53	0.21	0.42	0.4	0.48	0.021
M4						0.28	0.48	0.21	0.29	0.29	0.29	0.16	0.33	0.19	0.27	0.085	0.69	-0.18
HS2	0.2	0.6	0.17	0.83	0.29		0.17	-0.021	-0.2	0.51	0.26	0.17	0.075	-0.31	0.36	0.22	0.11	0.35
HS4	0.4	0.6	0.5	0.5	0.57			0.64	0.55	-0.029	0.33	0.36	0.75	0.27	0.31	0.51	0.34	0.18
HS3	0.4	0.2	0.33	0.33	0.29				0.67	-0.21	0.42	0.54	0.56	0.56	-0.13	0.7	0.093	0.4
HS8	0.2	0.2	0.33	0.33	0.43					-0.38	0.19	0.22	0.55	0.67	-0.027	0.57	0.013	0.11
HS9	0.4	0.4	0.33	0.33	0.14						0.6	0.22	-0.12	-0.52	0.48	-0.017	0.48	-0.16
HS1	0.4	0.4	0.17	0.5	0.14							0.54	0.18	-0.039	0.19	0.29	0.28	-0.036
HS5	1	0.4	0.67	0.67	0.43								0.33	-0.058	-0.063	0.41	0.16	0.39
HS12	0.2	0.2	0.17	0.33	0.14									0.45	0.31	0.75	0.0083	0.17
HS11	0.4	0.4	0.5	0.33	0.43										-0.27	0.47	-0.23	0.21
HS13	0.4	0.6	0.5	0.5	0.57											0.16	0.34	-0.33
HS10	0.4	0.6	0.5	0.5	0.29												-0.14	0.43
HS14	0.4	0.4	0.5	0.5	0.57													-0.39
HS15	0.2	0.2	0.17	0.33	0													
MEAN	0.38	0.4	0.37	0.46	0.33													

Table 31: Pearson's r and Precision (left below diagonal) for Article D.

	M0	M1	M2	M3	M4	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
M0		***	***	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	*	***	*	n.s.	n.s.	***	n.s.	n.s.	n.s.
M1			***	***	***	*	***	n.s.	n.s.	n.s.	*	n.s.	***	n.s.	***	*	n.s.	n.s.
M2				***	***	n.s.	*	*	n.s.	n.s.	*	*	*	n.s.	*	n.s.	***	n.s.
M3					***	n.s.	***	n.s.	*	n.s.	n.s.	n.s.	***	n.s.	*	*	*	n.s.
M4						n.s.	*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	***	n.s.
HS2	0.17	0.5	0.17	0.83	0.33		n.s.	n.s.	n.s.	*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
HS4	0.29	0.43	0.43	0.43	0.57			***	***	n.s.	n.s.	n.s.	***	n.s.	n.s.	*	n.s.	n.s.
HS3	0.4	0.2	0.4	0.4	0.4				***	n.s.	*	***	***	***	n.s.	***	n.s.	*
HS8	0.17	0.17	0.33	0.33	0.5					n.s.	n.s.	n.s.	***	***	n.s.	***	n.s.	n.s.
HS9	0.4	0.4	0.4	0.4	0.2						***	n.s.	n.s.	n.s.	*	n.s.	*	n.s.
HS1	0.4	0.4	0.2	0.6	0.2							***	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
HS5	0.71	0.29	0.57	0.57	0.43								n.s.	n.s.	n.s.	*	n.s.	n.s.
HS12	0.33	0.33	0.33	0.67	0.33									*	n.s.	***	n.s.	n.s.
HS11	0.33	0.33	0.5	0.33	0.5										n.s.	*	n.s.	n.s.
HS13	0.33	0.5	0.5	0.5	0.67											n.s.	n.s.	n.s.
HS10	0.29	0.43	0.43	0.43	0.29												n.s.	*
HS14	0.4	0.4	0.6	0.6	0.8													n.s.
HS15	0.17	0.17	0.17	0.33	0													
MEAN	0.34	0.35	0.39	0.49	0.4													

Table 32: Significance of Pearson's r and Recall (left below diagonal) for Article D.

	M0	M1	M2	M3	M4	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
M0		0.55	0.86	0.67	0.79	0.5	0.15	0.33	0.065	0.17	0.18	0.32	0.59	-0.023	0.57	0.48	0.25	-0.012
M1			0.59	0.28	0.44	0.21	0.16	-0.13	-0.07	0.078	0.35	0.092	-0.07	0.26	0.17	0.66	-0.14	0.45
M2				0.75	0.88	0.64	0.3	0.23	0.084	0.11	0.36	0.29	0.47	0.22	0.47	0.43	0.24	0.12
M3					0.78	0.5	-0.06	0.16	-0.18	-0.14	0.21	0.32	0.47	0.038	0.34	0.26	0.14	-0.22
M4						0.52	0.41	0.23	0.1	0.14	0.24	0.52	0.4	0.088	0.39	0.29	0.29	0.13
HS2	0.8	0.5	0.8	0.83	0.71		0.25	0.48	0.36	0.27	0.65	0.24	0.22	0.31	0.33	0.23	0.44	0.24
HS4	0.2	0.25	0.2	0.17	0.14			0.35	0.63	0.45	0.44	0.55	0.1	0.52	0.17	-0.047	0.58	0.72
HS3	0.4	0.25	0.4	0.33	0.29				0.31	0.51	0.35	0.54	0.47	-0.07	0.08	0.042	0.57	0
HS8	0.4	0.25	0.4	0.33	0.29					0.27	0.42	0.27	-0.023	0.55	0.23	-0.21	0.68	0.54
HS9	0.4	0.25	0.4	0.5	0.43						0.51	0.61	0.065	0.1	-0.1	-0.056	0.21	0.4
HS1	0.4	0.75	0.4	0.5	0.43							0.34	-0.061	0.57	0.12	0.31	0.24	0.64
HS5	0.2	0.25	0.2	0.33	0.29								0.28	0.078	0.055	-0.074	0.41	0.26
HS12	1	0.25	1	1	0.86									-0.028	0.65	0.033	0.39	-0.33
HS11	0.6	0.5	0.6	0.5	0.43										0.24	-0.059	0.47	0.72
HS13	0.6	0.25	0.6	0.67	0.57											0.4	0.19	0.04
HS10	0.4	0.75	0.4	0.33	0.43												-0.37	0.18
HS14	0.6	0.25	0.6	0.5	0.43													0.32
HS15	0.2	0.75	0.2	0.17	0.14													
MEAN	0.48	0.4	0.48	0.47	0.42													

Table 33: Pearson's r and Precision (left below diagonal) for Article E.

	M0	M1	M2	M3	M4	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
M0		***	***	***	***	*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	***	n.s.	***	*	n.s.	n.s.
M1			***	n.s.	*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	***	n.s.	*
M2				***	***	***	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.	*	*	n.s.	n.s.
M3					***	*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	n.s.	n.s.	n.s.
M4						*	*	n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
HS2	0.57	0.29	0.57	0.71	0.71		n.s.	*	n.s.	n.s.	***	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.
HS4	0.2	0.2	0.2	0.2	0.2			n.s.	***	*	*	***	n.s.	*	n.s.	n.s.	***	***
HS3	0.4	0.2	0.4	0.4	0.4				n.s.	*	n.s.	*	*	n.s.	n.s.	n.s.	***	n.s.
HS8	0.33	0.17	0.33	0.33	0.33					n.s.	*	n.s.	n.s.	***	n.s.	n.s.	***	*
HS9	0.4	0.2	0.4	0.6	0.6						*	***	n.s.	n.s.	n.s.	n.s.	n.s.	*
HS1	0.29	0.43	0.29	0.43	0.43							n.s.	n.s.	***	n.s.	n.s.	n.s.	***
HS5	0.2	0.2	0.2	0.4	0.4								n.s.	n.s.	n.s.	n.s.	*	n.s.
HS12	0.5	0.1	0.5	0.6	0.6									n.s.	***	n.s.	n.s.	n.s.
HS11	0.5	0.33	0.5	0.5	0.5										n.s.	n.s.	*	***
HS13	0.43	0.14	0.43	0.57	0.57											n.s.	n.s.	n.s.
HS10	0.29	0.43	0.29	0.29	0.43												n.s.	n.s.
HS14	0.5	0.17	0.5	0.5	0.5													n.s.
HS15	0.17	0.5	0.17	0.17	0.17													
MEAN	0.37	0.26	0.37	0.44	0.45													

Table 34: Significance of Pearson's r and Recall (left below diagonal) for Article E.

	M0	M1	M2	M3	M4	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
M0		0.68	0.93	0.74	0.94	0.54	0.46	0.61	0.31	0.67	0.58	0.61	0.63	0.11	0.62	0.65	0.39	0.51
M1			0.56	0.72	0.56	0.13	0.44	0.42	0.22	0.35	0.21	0.39	0.42	0.2	0.23	0.62	0.3	0.49
M2				0.75	0.9	0.57	0.52	0.56	0.31	0.65	0.47	0.58	0.61	0.11	0.6	0.56	0.39	0.53
M3					0.71	0.38	0.54	0.64	0.43	0.39	0.38	0.48	0.54	0.34	0.37	0.6	0.6	0.74
M4						0.6	0.41	0.67	0.23	0.75	0.69	0.66	0.58	0	0.65	0.68	0.29	0.49
HS2	0.5	0.33	0.5	0.43	0.5		0.6	0.43	0.095	0.65	0.54	0.49	0.4	0.33	0.86	0.42	0.21	0.37
HS4	0.67	0.33	0.67	0.57	0.67			0.31	0.13	0.24	0.24	0.52	0.4	0.36	0.66	0.32	0.17	0.35
HS3	0.5	0.5	0.5	0.43	0.5				0.35	0.53	0.73	0.53	0.41	0.21	0.48	0.45	0.4	0.67
HS8	0.5	0.33	0.5	0.57	0.5					0.12	0.044	0.38	0.48	0.66	0.021	0.37	0.72	0.16
HS9	0.83	0.5	0.83	0.43	0.83						0.69	0.52	0.35	0.027	0.56	0.66	0.15	0.2
HS1	0.83	0.33	0.83	0.57	0.83							0.34	0.09	0.037	0.47	0.38	0.22	0.28
HS5	0.67	0.67	0.67	0.43	0.67								0.8	0.07	0.69	0.54	0.071	0.33
HS12	0.17	0.5	0.17	0.43	0.17									0.12	0.61	0.38	0.22	0.44
HS11	0.33	0.33	0.33	0.43	0.33										0.035	0.19	0.62	0.12
HS13	0.67	0.5	0.67	0.57	0.67											0.42	0.034	0.44
HS10	0.67	0.83	0.67	0.57	0.67												0.31	0.35
HS14	0.5	0.33	0.5	0.57	0.5													0.46
HS15	0.67	0.67	0.67	0.71	0.67													
MEAN	0.58	0.47	0.58	0.52	0.58													

Table 35: Pearson's r and Precision (left below diagonal) for Article F.

	M0	M1	M2	M3	M4	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
M0		***	***	***	***	*	*	***	n.s.	***	***	***	***	n.s.	***	***	n.s.	*
M1			***	***	***	n.s.	*	*	n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	***	n.s.	*
M2				***	***	***	*	***	n.s.	***	*	***	***	n.s.	***	***	n.s.	*
M3					***	n.s.	*	***	*	n.s.	n.s.	*	*	n.s.	n.s.	***	***	***
M4						***	*	***	n.s.	***	***	***	***	n.s.	***	***	n.s.	*
HS2	0.6	0.4	0.6	0.6	0.6		***	*	n.s.	***	*	*	*	n.s.	***	*	n.s.	n.s.
HS4	0.67	0.33	0.67	0.67	0.67			n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	***	n.s.	n.s.	n.s.
HS3	0.6	0.6	0.6	0.6	0.6				n.s.	*	***	*	*	n.s.	*	*	*	***
HS8	0.5	0.33	0.5	0.67	0.5					n.s.	n.s.	n.s.	*	***	n.s.	n.s.	***	n.s.
HS9	0.71	0.43	0.71	0.43	0.71						***	*	n.s.	n.s.	***	***	n.s.	n.s.
HS1	0.83	0.33	0.83	0.67	0.83							n.s.	n.s.	n.s.	*	n.s.	n.s.	n.s.
HS5	0.67	0.67	0.67	0.5	0.67								***	n.s.	***	*	n.s.	n.s.
HS12	0.2	0.6	0.2	0.6	0.2									n.s.	***	n.s.	n.s.	*
HS11	0.33	0.33	0.33	0.5	0.33										n.s.	n.s.	***	n.s.
HS13	0.67	0.5	0.67	0.67	0.67											*	n.s.	*
HS10	0.67	0.83	0.67	0.67	0.67												n.s.	n.s.
HS14	0.5	0.33	0.5	0.67	0.5													*
HS15	0.67	0.67	0.67	0.83	0.67													
MEAN	0.59	0.49	0.59	0.62	0.59													

Table 36: Significance of Pearson's r and Recall (left below diagonal) for Article F.

	M0	M1	M2	M3	M4	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
M0		0.46	0.88	0.45	0.67	0.25	0.26	0.27	0.1	0.35	0.33	0.15	0.26	0.036	0.26	0.24	0.27	0.062
M1			0.44	0.56	0.42	0.29	0.27	0.33	0.22	0.22	0.32	0.24	0.34	0.21	0.37	0.59	0.25	0.27
M2				0.53	0.78	0.26	0.31	0.29	0.14	0.34	0.35	0.1	0.3	0.078	0.29	0.22	0.31	0.12
M3					0.61	0.41	0.33	0.39	0.29	0.17	0.33	0.28	0.51	0.24	0.32	0.35	0.36	0.23
M4						0.29	0.28	0.34	0.19	0.29	0.34	0.18	0.29	0.063	0.31	0.2	0.33	0.13
HS2	0.48	0.45	0.47	0.56	0.47		0.42	0.31	0.3	0.39	0.44	0.3	0.23	0.41	0.51	0.35	0.35	0.37
HS4	0.36	0.33	0.39	0.41	0.4			0.37	0.48	0.46	0.47	0.41	0.35	0.6	0.43	0.36	0.42	0.51
HS3	0.39	0.45	0.39	0.44	0.37				0.42	0.26	0.43	0.49	0.41	0.23	0.31	0.42	0.37	0.31
HS8	0.36	0.36	0.39	0.37	0.37					0.23	0.28	0.37	0.31	0.57	0.34	0.37	0.45	0.3
HS9	0.42	0.36	0.42	0.44	0.37						0.6	0.37	0.18	0.17	0.35	0.25	0.43	0.24
HS1	0.42	0.45	0.36	0.49	0.4							0.35	0.31	0.32	0.4	0.37	0.37	0.32
HS5	0.45	0.42	0.42	0.44	0.4								0.47	0.26	0.36	0.26	0.32	0.3
HS12	0.36	0.33	0.36	0.51	0.33									0.27	0.44	0.38	0.23	0.15
HS11	0.42	0.45	0.47	0.46	0.47										0.36	0.44	0.34	0.48
HS13	0.39	0.48	0.42	0.51	0.44											0.5	0.29	0.22
HS10	0.45	0.76	0.44	0.46	0.4												0.16	0.34
HS14	0.45	0.36	0.47	0.51	0.47													0.28
HS15	0.33	0.48	0.33	0.46	0.3													
MEAN	0.41	0.44	0.41	0.47	0.4													

Table 37: Pearson's r and Precision (left below diagonal) for all Articles (first sentences are included).

	M0	M1	M2	M3	M4	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
M0		***	***	***	***	***	***	***	n.s.	***	***	n.s.	***	n.s.	***	***	***	n.s.
M1			***	***	***	***	***	***	*	*	***	***	***	*	***	***	***	***
M2				***	***	***	***	***	n.s.	***	***	n.s.	***	n.s.	***	*	***	n.s.
M3					***	***	***	***	***	*	***	***	***	***	***	***	***	***
M4						***	***	***	*	***	***	*	***	n.s.	***	*	***	n.s.
HS2	0.44	0.42	0.47	0.64	0.56		***	***	***	***	***	***	***	***	***	***	***	***
HS4	0.33	0.31	0.39	0.47	0.47			***	***	***	***	***	***	***	***	***	***	***
HS3	0.39	0.45	0.42	0.55	0.48				***	***	***	***	***	*	***	***	***	***
HS8	0.35	0.35	0.41	0.44	0.47					*	***	***	***	***	***	***	***	***
HS9	0.39	0.33	0.42	0.5	0.44						***	***	*	*	***	***	***	***
HS1	0.37	0.39	0.34	0.53	0.45							***	***	***	***	***	***	***
HS5	0.44	0.41	0.44	0.53	0.5								***	***	***	***	***	***
HS12	0.38	0.34	0.41	0.66	0.44									***	***	***	***	n.s.
HS11	0.37	0.39	0.45	0.5	0.53										***	***	***	***
HS13	0.36	0.44	0.42	0.58	0.53											***	***	*
HS10	0.37	0.61	0.39	0.46	0.41												n.s.	***
HS14	0.44	0.35	0.5	0.62	0.59													***
HS15	0.31	0.44	0.33	0.53	0.36													
MEAN	0.38	0.4	0.41	0.54	0.48													

Table 38: Significance of Pearson's r and Recall (left below diagonal) for all Articles (first sentences are included).

	M0	M1	M2	M3	M4	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
M0		0.25	0.85	0.31	0.59	0.15	0.15	0.16	-0.0085	0.29	0.2	0.038	0.1	-0.072	0.1	0.091	0.21	0.036
M1			0.22	0.4	0.23	0.18	0.13	0.19	0.088	0.096	0.13	0.1	0.12	0.094	0.17	0.49	0.18	0.29
M2				0.41	0.73	0.17	0.21	0.19	0.035	0.27	0.23	-0.019	0.15	-0.024	0.13	0.062	0.26	0.092
M3					0.52	0.35	0.24	0.3	0.21	0.076	0.2	0.18	0.42	0.17	0.17	0.23	0.31	0.22
M4						0.2	0.18	0.25	0.1	0.23	0.22	0.077	0.15	-0.029	0.17	0.057	0.29	0.1
HS2	0.39	0.37	0.39	0.51	0.39		0.38	0.25	0.26	0.35	0.38	0.24	0.15	0.38	0.47	0.28	0.32	0.37
HS4	0.29	0.26	0.32	0.37	0.34			0.32	0.45	0.43	0.41	0.37	0.28	0.58	0.37	0.3	0.39	0.51
HS3	0.29	0.33	0.29	0.34	0.29				0.38	0.19	0.36	0.44	0.33	0.17	0.2	0.34	0.34	0.31
HS8	0.25	0.22	0.29	0.26	0.29					0.18	0.22	0.32	0.23	0.54	0.27	0.31	0.43	0.3
HS9	0.32	0.22	0.32	0.34	0.29						0.57	0.33	0.092	0.12	0.28	0.17	0.4	0.23
HS1	0.32	0.33	0.26	0.4	0.32							0.28	0.19	0.27	0.29	0.27	0.33	0.32
HS5	0.36	0.3	0.32	0.34	0.32								0.41	0.21	0.27	0.17	0.28	0.3
HS12	0.25	0.19	0.26	0.43	0.24									0.2	0.32	0.26	0.17	0.14
HS11	0.32	0.33	0.39	0.37	0.39										0.3	0.4	0.32	0.48
HS13	0.29	0.37	0.32	0.43	0.37											0.4	0.24	0.22
HS10	0.36	0.7	0.35	0.37	0.32												0.1	0.34
HS14	0.43	0.33	0.45	0.51	0.45													0.27
HS15	0.29	0.48	0.29	0.46	0.26													
MEAN	0.32	0.34	0.33	0.4	0.33													

Table 39: Pearson's r and Precision (left below diagonal) for all Articles (first sentences are excluded).

	M0	M1	M2	M3	M4	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
M0		***	***	***	***	n.s.	n.s.	n.s.	n.s.	***	*	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.
M1			*	***	***	*	n.s.	*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	*	***	*	***
M2				***	***	*	*	*	n.s.	***	*	n.s.	n.s.	n.s.	n.s.	n.s.	***	n.s.
M3					***	***	***	***	*	n.s.	*	*	***	*	*	***	***	*
M4						*	*	***	n.s.	*	*	n.s.	n.s.	n.s.	*	n.s.	***	n.s.
HS2	0.35	0.32	0.39	0.58	0.48		***	***	***	***	***	***	n.s.	***	***	***	***	***
HS4	0.25	0.22	0.31	0.41	0.41			***	***	***	***	***	***	***	***	***	***	***
HS3	0.3	0.33	0.33	0.44	0.41				***	*	***	***	***	*	*	***	***	***
HS8	0.25	0.21	0.32	0.32	0.39					*	*	***	***	***	***	***	***	***
HS9	0.3	0.2	0.33	0.4	0.37						***	***	n.s.	n.s.	***	*	***	***
HS1	0.28	0.28	0.25	0.44	0.38							***	*	***	***	***	***	***
HS5	0.36	0.29	0.36	0.43	0.43								***	*	***	*	***	***
HS12	0.27	0.19	0.31	0.58	0.35									*	***	***	*	n.s.
HS11	0.28	0.28	0.38	0.41	0.47										***	***	***	***
HS13	0.27	0.33	0.33	0.5	0.47											***	***	*
HS10	0.29	0.54	0.31	0.37	0.34												n.s.	***
HS14	0.39	0.29	0.45	0.58	0.55													***
HS15	0.24	0.39	0.27	0.48	0.3													
MEAN	0.29	0.3	0.33	0.46	0.41													

Table 40: Significance of Pearson's r and Recall (left below diagonal) for all Articles (first sentences are excluded).

	M0	M1	M2	M3	M4	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
M0		***	***	***	***	n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
M1			***	***	***	*	n.s.	***	n.s.	*	***	n.s.	*	n.s.	***	***	***	*
M2				*	***	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.
M3					***	***	n.s.	*	n.s.	n.s.	n.s.	n.s.	***	n.s.	n.s.	n.s.	***	n.s.
M4						*	n.s.	***	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	***	n.s.
HS2							***	n.s.	n.s.	n.s.	*	n.s.	***	*	***	n.s.	*	n.s.
HS4								***	***	***	***	***	n.s.	***	***	n.s.	***	***
HS3									*	***	n.s.	***	*	n.s.	n.s.	*	***	***
HS8										n.s.	*	***	***	*	*	n.s.	***	n.s.
HS9											***	***	n.s.	n.s.	n.s.	n.s.	***	n.s.
HS1												*	*	n.s.	***	***	***	*
HS5													n.s.	n.s.	n.s.	n.s.	*	n.s.
HS12														n.s.	***	n.s.	n.s.	n.s.
HS11															n.s.	***	n.s.	***
HS13																***	*	n.s.
HS10																	n.s.	*
HS14																		n.s.

Table 41: Significance of Chi Squares: Sentence Scores: 5/4 vs. 3/2/1 (first sentences are included).

	M0	M1	M2	M3	M4	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
M0		n.s.	***	n.s.	***	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
M1			n.s.	***	*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	*	*	*
M2				n.s.	***	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
M3					*	*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	***	n.s.	n.s.	n.s.	*	n.s.
M4						n.s.	n.s.	*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.
HS2							*	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.	*	n.s.	n.s.	n.s.
HS4								***	*	***	*	***	n.s.	*	***	n.s.	***	***
HS3									n.s.	***	n.s.	***	n.s.	n.s.	n.s.	n.s.	***	***
HS8										n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	***	*
HS9											***	*	n.s.	n.s.	n.s.	n.s.	***	n.s.
HS1												n.s.	n.s.	n.s.	n.s.	***	n.s.	*
HS5													n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
HS12														n.s.	***	n.s.	n.s.	n.s.
HS11															n.s.	*	n.s.	***
HS13																***	n.s.	n.s.
HS10																	n.s.	*
HS14																		n.s.

Table 42: Significance of Chi Squares: Sentence Scores: 5/4 vs. 3/2/1 (first sentences are excluded).

	M0	M1	M2	M3	M4	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
M0		n.s.	***	***	***	*	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	n.s.	n.s.	*	n.s.
M1			n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	*	***	n.s.	*
M2				***	***	*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	***	n.s.
M3					***	***	n.s.	*	n.s.	*	*	*	***	*	***	n.s.	***	*
M4						*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	*	*	n.s.	***	n.s.
HS2							***	*	*	***	***	n.s.	***	*	***	*	***	n.s.
HS4								***	***	*	***	***	n.s.	***	***	*	n.s.	***
HS3									***	***	***	***	***	***	n.s.	***	*	***
HS8										***	*	*	***	***	n.s.	***	*	n.s.
HS9											***	***	n.s.	*	***	n.s.	***	n.s.
HS1												***	n.s.	***	*	*	n.s.	***
HS5													***	***	***	*	n.s.	***
HS12														***	***	*	n.s.	n.s.
HS11															*	*	*	***
HS13																***	n.s.	n.s.
HS10																	n.s.	***
HS14																		*

Table 43: Significance of Chi Squares: Selected Sentences (first sentences are included).

	M0	M1	M2	M3	M4	HS2	HS4	HS3	HS8	HS9	HS1	HS5	HS12	HS11	HS13	HS10	HS14	HS15
M0		n.s.	***	***	***	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
M1			n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	***	n.s.	*
M2				***	***	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.
M3					***	***	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	***	n.s.	*	n.s.	***	*
M4						n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	***	n.s.
HS2							*	n.s.	n.s.	***	*	n.s.	*	n.s.	*	n.s.	*	n.s.
HS4								*	***	n.s.	***	*	n.s.	***	*	n.s.	n.s.	***
HS3									***	n.s.	n.s.	***	*	***	n.s.	*	n.s.	*
HS8										n.s.	n.s.	n.s.	n.s.	***	n.s.	*	*	n.s.
HS9											***	*	n.s.	n.s.	n.s.	n.s.	*	n.s.
HS1												***	n.s.	*	n.s.	n.s.	n.s.	*
HS5													n.s.	n.s.	n.s.	n.s.	n.s.	*
HS12														n.s.	***	n.s.	n.s.	n.s.
HS11															n.s.	n.s.	n.s.	***
HS13																n.s.	n.s.	n.s.
HS10																	n.s.	***
HS14																		n.s.

Table 44: Significance of Chi Squares: Selected Sentences (first sentences are excluded).

E README: Location, Description, and Usage of the Executables, Sources and Various Other Files

```
*****************************************************
This file: /import/usersA/klaus/sumproj/lookat/README
Last edit: 5 August 95
*****************************************************


This directory contains the following files:

   3 -rw-r--r--  1 klaus        3014 Jul 13 11:26 README

   4 -rw-rw-rw-  1 klaus        3996 Aug  5 15:27 docabstr.info
  38 -rw-rw-rw-  1 klaus       38000 Aug  5 15:27 precis.cc
1064 -rwxrwxrwx  1 klaus     1081344 Aug  5 15:26 doprecis*

  11 -rw-r--r--  1 klaus       10883 Aug  5 15:34 art.a
  13 -rw-r--r--  1 klaus       12793 Aug  5 15:34 art.b
  12 -rw-r--r--  1 klaus       11898 Aug  5 15:34 art.c
  12 -rw-r--r--  1 klaus       11627 Aug  5 15:34 art.d
  12 -rw-r--r--  1 klaus       11836 Aug  5 15:34 art.e
  12 -rw-r--r--  1 klaus       11715 Aug  5 15:34 art.f
 120 -rw-r--r--  1 klaus      112919 Aug  5 15:34 all.out

  47 -rw-rw-rw-  1 klaus       47282 Aug  5 15:29 stat.cc
 488 -rwxrwxrwx  1 klaus      491520 Aug  5 15:29 dostat*

  13 -rw-r--r--  1 klaus       12368 Aug  5 15:37 res.12-14
  11 -rw-r--r--  1 klaus       11166 Aug  5 15:37 res.4-6
  12 -rw-r--r--  1 klaus       11769 Aug  5 15:37 res.8-10
  12 -rw-r--r--  1 klaus       11295 Aug  5 15:37 res.cw5-11
   7 -rw-r--r--  1 klaus        6220 Aug  5 15:29 res.humans.only

 184 -rw-r--r--  1 klaus      173893 Aug  5 15:37 r.12-14
 176 -rw-r--r--  1 klaus      171336 Aug  5 15:37 r.4-6
 184 -rw-r--r--  1 klaus      173528 Aug  5 15:37 r.8-10
 176 -rw-r--r--  1 klaus      170393 Aug  5 15:37 r.cw5-11
  64 -rw-r--r--  1 klaus       65013 Aug  5 15:38 rho


Explanation:
************

(1) Abstracting:
****************

(1.1) executable: doprecis* (corresponding source: precis.cc)

 Use program doprecis* and specify an SGML-file, which is at least
marked up for words <W> and sentences <S> and contains one article
<REUT>, as input file, e.g.:

scott% doprecis* art.a
```

```
This will produce an output-file with suffix .out which contains
tables about word- and sentence-scores and -rankings and abstracts
produced by 5 different sentence weighting methods.

(1.2) parameter-file: docabstr.info

Various parameters for doprecis* can be set in this file. See the
description in the file itself and also in the MSc dissertation,
section 2.

(1.3) input files: art.a ... art.f (6 articles, those from the experiment)

(1.4) sample output file:  all.out (6 articles, those from the experiment)



(2) Statistics:
***************

(2.1) executable: dostat* (corresponding source: stat.cc)

 Use program dostat* for a statistics of sentence scores and selected
sentences of various subjects (humans/machine methods). One argument
is required: it specifies the input file.
 Redirect the output to a file; it is advisable to redirect the error
output as well, e.g.:

scott% dostat* res.4-6 >& r.4-6

The statistics contain:
  - mean scores and standard deviations for all sentences
  - correlation of scores between all subjects for all articles
  - chi square evaluations

(2.2) sample files: res.* (results of various test runs of the
      doprecis* program, combined with the results of the experiment)

(2.3) result files: r.*, rho (results for human subjects only, using res.humans.only)


************ End of File: README ************************************
```