Selective Text Traversal

Gerard Salton* and Amit Singhal

Abstract

In information retrieval and text processing, the size of the available databases and text collections has grown enormously in the past few years, and users find it difficult to cope with the amount of text that is potentially of interest. As a result, there is growing interest in approaches capable of breaking down large texts into smaller constituent units, and in using short text passages for storage and retrieval purposes. Methods are described in this study for identifying important text passages in large texts, and suggestions are made for implementing useful text reading and text traversal strategies that provide selective text access in accordance with user needs. Some evaluation results for selective text access are appended.

1 Background on Passage Retrieval

The idea of introducing passage retrieval techniques in information retrieval as a replacement for full-text processing only has been discussed in the literature early on. [1] As the size of available full-text databases has grown over the years, the interest in the treatment of individual text passages has grown considerably, and passages are now used both as independent retrieval units, and also as enhancements for the retrieval of full-text information.

Considering first the use of text passages as separate retrieval units, it is well known that large, heterogeneous documents may be difficult to retrieve in response to narrow, specific queries. Specifically, in many vector processing environments, document representations are usually normalized for document length to avoid

*Department of Computer Science, Cornell University, Ithaca, NY 14853-7501. This study was supported in part by the National Science Foundation under grant IRI 9300124.

favoring the larger documents over shorter ones. In such circumstances, the coincidence between a long text that covers many topics and a more specific query will be small, and the retrieval status value for such a text will be too low for retrieval to occur. When such a text is broken down into individual passages each with homogeneous subject coverage, the retrieval value of particular passages that directly match a given query topic will be much higher than that of the full text. This explains the success of the mixed retrieval strategy implemented in the Smart retrieval environment which makes it possible to retrieve particular text elements, such as text paragraphs or text sections, whenever these text elements provide a better match with the query than the full-text. [2,3]

In addition to direct retrieval of text passages, text excerpts are also usable as enhancements to full-text processing in various ways. In particular, in a global text matching environment where full document texts are compared with query texts, it is often difficult to eliminate false retrievals produced by matching text identifiers that carry distinct meanings. In such circumstances, it is necessary to look at the local environments in which the terms are used in the respective texts. For example, if a term such as "base" occurs in a text sentence, or paragraph, with other terms such as hit, run, bunt, strike, and so on, the subject is likely to be baseball. By using local text passages to verify the context in which matching query terms occur, it becomes possible to reject items that do not exhibit properly matching local contexts with the query.

In the Smart system, a dual global plus local text comparison operation is carried out, and documents with appropriate global query similarity are retrieved only if a sufficiently large similarity also exists between the query and a local text passage. [4,5] In recent text retrieval experiments, a number of research groups have used such passage-level evidence with considerable success to control or enhance text retrieval, based on local contexts such as variable-length text passages (paragraphs or sentences) [6-10], or short text windows of fixed-length [11-12].

2 Text Structure Analysis

In the preceding section, text passages were introduced as useful tools in information retrieval. In a modern full-text processing environment, a wide array of local processing operations may be considered based first on a decomposition of full texts into semantically homogeneous subunits, followed by a re-assembly of certain units as required for particular applications. For example, a text summary might be constructed by locating important text excerpts and then repackaging these excerpts into a complete summarizing unit. In this context, general automated text writing and reading environments have been discussed that provide flexible text manipulation operations for many user classes. [13-16] Methods have also been introduced for transforming standard written texts into so-called hypertexts, consisting of linked text pieces. Typically, a graph structure is used to represent one or more texts in which the nodes, or vertices, represent pieces of text, and the branches or links between nodes identify certain kinds of relationships between the corresponding text pairs. [17-21]

In conventional hypertext environments, the branches are normally assigned by experts who are familiar with the subject matter and know what kind of text relationships are important in a particular application. Links can, however, also be assigned automatically, for example by using the formal structural characteristics of texts to relate individual sentences to particular paragraphs containing them, or individual paragraphs to sections, and sections to chapters. In that case, text linking is used to represent the normal physical text structure. Links related to the text composition can also be placed automatically, for example, by linking the various occurrences of individual words or phrases as they occur in a text.

In a vector processing environment, such as the Smart system, a text relationship map can be built based on the computed similarity between all text pairs. Typically, a threshold value is picked for the text similarity, known as the compare similarity (csim) value, and all pairs of texts whose similarity exceeds the threshold are joined by a link. [22-24] A typical text relationship map in the general area of nuclear energy is shown in Figure 1 for six encyclopedia articles from the New Funk and Wagnall's encyclopedia. A similarity threshold of 0.10 is used in Figure 1. The map shows that a substantial similarity above threshold exists between the item on Nuclear Weapons (document 17016) and both Nuclear Fission (document 8907) and Thermonuclear

Fission (document 22387). However, no direct relationship exists between the latter two items dealing with nuclear fusion and fission, respectively.

Text relationship maps are also usable to represent similarities between text passages, rather than full texts only. Figure 2 presents a sample paragraph relationship map where the nodes are text paragraphs, and the branches are paragraph similarities. The graph of Figure 2 shows that the relationship between the text on Nuclear Weapons (item 17016) and the other two items on Nuclear Fusion and Fission is established by paragraphs 3 of these texts (8907.p3 and 22387.p3, respectively) describing the use of nuclear fission and fusion for nuclear weapons development. In general, the text relationship display is thus useful for the identification of related text passages covering particular homogeneous topic areas.

3 Generation of Text Themes

A text theme may be defined as an area of concentration in a running text consisting of sets of related text excerpts covering a common related subject area. In a text relationship map, a theme is often characterized by a collection of text nodes with substantial numbers of links between them. Attempts have been made in the past to identify text themes by locating shifts in topic coverage in normal running text. [25]

The approach proposed here for text theme identification is based on an initial recognition of triangles in the text relationship maps, followed by a merging of triangles whose descriptions are sufficiently similar. [26-27] Each triangle identifies a set of three mutually related text items, and is represented by a term vector (the centroid) computed as the average of the three component texts. When the similarity between two triangle centroids is sufficiently large, the triangles are merged into a larger-order structure, and each set of merged triangles is then associated with a text theme.

Consider, as an example, the paragraph relationship map of Figure 3 representing paragraph similarities for document 17016 on Nuclear Weapons. Three themes are obtained following the recognition and merging of triangles:

1. Theme 1 consisting of paragraphs 11, 13, 23, and 26 roughly entitled, “A-bomb and Related

Topics". The main terms characterizing this theme are atomic fission, uranium isotope, atomic nucleus, and chain reaction.

2. Theme 2 consisting of paragraphs 5, 29, 30, and 35 entitled, "H-bomb and related topics", characterized by terms thermonuclear reactions, fusion of light elements, hydrogen and helium.
3. Theme 3 consisting of paragraphs 33, 35, 48, 49, 50 and 54 dealing with nuclear radiation and related topics. The main terms are radioactive fallout, thermonuclear contamination, radioactive debris and neutron bomb.

Two main parameters central to the text theme generation process are the link density of the text relationship map, that is, the ratio of the number of links on the map to the number of nodes, and the similarity threshold controlling the triangle merging process. The link density depends on the threshold in the similarity value (the csim value) used to construct the map – the lower the csim value, the more links will appear on the map, and hence the higher the link density. The other parameter, represents the threshold controlling the triangle merging, known as the tsim value. When the centroid similarity for two triangles exceeds the tsim value, the triangles are merged. An outline of the theme generation process appears in Table 1. To minimize the overlap between themes (the number of paragraphs in common between them), and simplify the presentation, a low density specification of 0.75 is used for the map of Figure 3, corresponding to a csim value of 0.40. A higher density specification of 1.50 would have doubled the number of allowed links from 27 in Figure 3 to 54.

Since each theme represents a highlight in the subject coverage of a document, the text excerpts forming the themes might be used to provide specific answers to user queries in a retrieval environment. Analogously, the full set of themes might serve for text summarization purposes. For the example of Figure 3, the theme set consists of 13 distinct paragraphs compared with the 54 paragraphs in the full text. A replacement of the full document texts by theme information is thus likely to increase substantially the efficiency of many text processing operations.

4 Text Traversal Paths

The text relationship maps used for theme generation can also represent other aspects of text structure. In particular, the maps are usable to control text traversal strategies by identifying important text excerpts, and then associating an ordered set of excerpts with a selective text traversal specification. Rapid text traversal and text skimming operations can provide a browsing environment that saves user time and effort. In addition, selective text traversal may also enhance the effectiveness of text usage, because the text traversal paths often confine themselves to particular subject areas covered in the document texts while excluding much of the extraneous information.

In choosing text traversal strategies, various criteria suggest themselves such as:

- the coherence of the traversal path, that is, the degree to which the chosen text excerpts hang together properly;
- the comprehensiveness of the path, that is, the degree to which most aspects of text content are included in the path;
- the length of the path as measured by the number of included text excerpts;
- the location in the full text of the extracted material (for example, initial paths include much of the introductory material, whereas terminal paths include concluding material);
- the centrality of the nodes included in a path, measured by the number of incident links attached to each node ("bushy" nodes with many links to other nodes may be assumed to be central to text content).

In general, the use of continuous paths where each included excerpt is related to the subsequent excerpt in the path order is most likely to insure text coherence. A depth-first path, produced by taking an initial node and iteratively moving to the next most similar node, is a typical example of such a continuous path. Depth-first paths are often not comprehensive, because attention may be concentrated on particular aspects of text content. If comprehensiveness is important, the "bushy" paths that include most of the central, heavily linked nodes are often preferred. A special path type known as a tour, includes only text excerpts included in one or more of the themes. The longer tours tend to produce comprehensive text traversal patterns.

In defining a traversal order, the nodes included in a path are listed in chronological text order, that is, no backtracking is allowed. This is of consequence especially for depth-first paths where the node most related to some node x might precede x in the normal text order. The path length for depth-first paths is not specifiable in advance because the length will depend on the connection pattern of the relationship maps. However, a desired length can be specified for other paths as a percentage of nodes included in the path, compared with the total length of the document. To construct an x percent path, the compare threshold (csim value) is raised, rendering the path more sparse. Low percentage paths include only the best nodes, that is, those with the most highly-weighted links, or with the greatest number of incident lines. Longer paths will include additional, less compelling nodes. Definitions for various path types are included in Table 2.

Some typical 5 percent paths are shown in Figure 4 for sample document 17016 on Nuclear Weapons. The graph of Figure 4 is based on a link density of 2.0, corresponding to a compare similarity of 0.28. Excerpts from the paragraphs included in each path of Figure 4 are reproduced in Table 3. As the examples show, the 5 percent initial path contains the main definition and an explanation of the scope of nuclear weapons. The 5 percent bushy path consists of one paragraph dealing with nuclear fission, and another dealing with nuclear fallout. The 5 percent tour, on the other hand, is entirely confined to excerpts from a theme (Theme 1 covering nuclear fission and the A-bomb). This is true also of the depth-first path for document 17016 starting at node p5. That path, not shown in Table 3, covers nodes 5, 29, 30, 33, 35, 48, and 54, and the subject matter is confined to Themes 2 and 3, but excludes the material of Theme 1.

Paths can be automatically restricted to particular subject areas by confining the text excerpts to particular areas of the document text. This is true, for example, of initial paths which cover the introductory material, and correspondingly of terminal paths. It is also true of paths restricted to excerpts within particular themes. An example of paths within themes is shown in Figure 5 for the sample document used earlier (number 17016). Figure 5 shows some depth-first paths within themes, and the corresponding text excerpts of Table 4 demonstrate that those paths indeed cover the essence of the three main subjects discussed in the document (nuclear fission, nuclear fusion, and radioactive fallout).

The sample paths included in this study are indicative of the variety of traversal orders that may prove useful in various circumstances. In each case, a successful path includes the most useful, or most important nodes, and rejects much of the extraneous materials. If efficiency is important, short paths should be used. For coherence, depth-first paths are preferred over tours and bushy paths. On the other hand if comprehensiveness is wanted, the bushy paths and the longer tours are best, because these paths include the important nodes only. A summary of path characteristics appears in Table 5.

5 Path Evaluation

The evaluation of passage retrieval techniques is fraught with difficulties because relevance assessments of documents with respect to queries are normally available only for full-text items and not for document passages. That is, when a text passage is retrieved in response to a query, no reliable information is provided about the usefulness of the retrieved item. In the present context, a somewhat complicated procedure is used for evaluation, which consists in first replacing each of 70 original search requests by a known relevant document with respect to each query. Each of these known relevant documents is then used instead of the given query to search a collection of some 46,000 documents (the Federal Register collection that forms a part of the well-known TREC collections [28]). The evaluation then consists in comparing the retrieval performance (11-point average precision over 70 searches) for the full documents used as queries, compared with the performance of various extracts from these documents used as queries. In all cases, the (somewhat dubious) assumption is made that the relevance assessments available for the original queries are valid also for the relevant documents; or their extracts when these are used as queries. The relevance assessments available in these circumstances will be satisfactory if the relevant papers chosen for query purposes cover the subject matter of the original queries reasonably closely.

Figures 6 and 7 contain differences in average (11-point) precision for the retrieval performance of various text extracts minus the retrieval performance of the full document texts (labeled 100 percent extract in the figures). Figure 6 covers the performance of initial paths, bushy paths, and tours. In that case, the path length can be controlled as previously explained. The Figure shows that for very short path lengths, the bushy paths are preferred, because they include only important paragraphs and often cover several main aspects of a document. However, even short bushy paths show a deficit of about 10 percent compared with the full texts. The shorter initial paths are not good because in the Federal Register, the documents often start with sets of addresses, or sets of references that do not directly relate to text content.

The main conclusion derivable from Figure 6 is that beyond a 20 percent length, that is, for extracts consisting of at least 20 percent of the full texts, the deficit for the extracts compared with full-length texts is normally less than 5 percent. For tours of length greater than 25 percent, the performance of the tours may in fact be better than that of the full text items. That is, sufficiently long tours, including excerpts from the various themes, tend to cover the document subject matter better than the full text items because the less important document components (such as addresses, signatures, references, etc.) that may decrease performance for the full texts are absent from the tours.

The performance of depth-first paths in the full texts, as well as depth-first paths confined to the largest theme - the theme with the largest number of components – is outlined in Figure 7. Since depth-first paths do not have a definite lengthy, each path is represented by a single point with an abscissa corresponding to the particular path length, and an ordinate giving the performance standard.

It is clear from Figure 7 that no substantial difference is visible between the full – depth-first paths and the theme 1 – depth-first paths. This is not surprising because even full depth-first extracts are normally limited to covering only a single theme. In both cases, the performance is clustered around the zero-line, where the performance of the text extract is the same as that of the full text. The data of Figure 7 also show that somewhat more point appear below the zero line than above it, indicating that a deficit arises for the depth-first paths compared with full texts more often than the reverse. The actual data for the 70 full-document depth-first paths show that the performance is equal to that of full texts for 15 queries, better for 20, and worse for 35. For the theme 1 – depth-first paths the corresponding figures are 18 equal, 23 better and 29 worse. Once again most of the performance differences are small. One concludes that only small losses arise when full document texts are replaced by suitably chosen text extracts, including theme information (tours), central text nodes (bushy paths), and continuous paths that start with an important text node (depth-first).

6 References

1. J. O'Connor, Answer Passage Retrieval by Text Searching, *Journal of the American Society for Information Science*, 32:4, July 1980, 227-239.

2. G. Salton, J. Allan and C. Buckley, Automatic Structuring and Retrieval of Large Text Files, *Communications of the ACM*, 37:2, February 1994, 97-108.

3. G. Salton, J. Allan and C. Buckley, Approaches to Passage Retrieval in Full Text Information Systems, Proc. SIGIR-93, *Association for Computing Machinery*, New York, June 1993, 49-58.

4. G. Salton and C. Buckley, Global Text Matching for Information Retrieval, Science, 253, 30 August 1991, 1012-1015.

5. G. Salton and C. Buckley, Automatic Text Structuring and Retrieval – Experiments in Automatic Encyclopedia Searching, Proc. SIGIR Conference '91, *Association for Computing Machinery*, New York, 21-30.

6. K.L. Kwok, L. Grunfeld, and D.D. Lewis, TREC 3 Ad-Hoc, Routing Retrieval, and Thresholding Experiments Using PIRCS, Text Retrieval Conference, Washington, DC, November 1994.

7. D. Kraus, E. Mittendorf, and P. Schauble, Improving a Basic Retrieval Method by Links and Passage Level Evidence, Text Retrieval Conference, Washington, DC, November 1994.

8. S.E. Robertson, S. Walton, S. Jones, M.M. Hancock-Beaulieu, and M. Gatford, Okapi at TREC 3, Text Retreival Conference, Washington, DC, November 1994.

9. R. Wilkinson, Effective Retrieval of Structured Documents, Proc. SIGIR '94, Springer Verlag, Berlin, 311-317.

10. E. Mittendorf and P. Schauble, Document and Passage Retrieval Based on Hidden Markov Models, Proc. SIGIR '94, Springer Verlag, Berlin, 318-327.

11. Chris Buckley, G. Salton, J. Allan and A. Singhal, Query Expansion Using Smart: TREC 3, Text Retrieval Conference, Washington, DC, November 1994.

12. J.P. Callan, Passage-Level Evidence in Document Retrieval, Proc. SIGIR '94, Springer Verlag, Berlin, 301-310.

13. J.D. Bolter, The Writing Space: The Computer, Hypertext, and the History of Writing, Lawrence Erlbaum, Hillsdale, N.J., 1991.

14. P. Delaney and G.P. Landow, eds., Hypermedia and Literary Studies, MIT Press, Cambridge MA, 1991.

15. S.P. DeRose, Biblical Studies and Hypertext, in P. Delany and G.P. Landow, (see reference [14]).

16. G.P. Landow, Hypertext in Literary Education, Criteria and Scholarship, *Computers and the Humanities 23,* 178-198, 1989.

17. R.H. Trigg and M. Weiser, Textnet: A Network-Based Approach to Text Handling, *Transactions on Office Information Systems*, 4:1, January 1986, 1-23.

18. P. Gloor, Cybermap: Yet Another Way of Navigating Hyperspace, Proc. Hypertext '91, *Association for Computing Machinery*, New York, December 1991, 107-121.

19. R.S. Gilyarevskii and M.M. Subbotin, Russian Experience in Hypertext, Automatic Computing of Coherent Texts, Journal of the Am. Soc. of Information Science, 44:4, 1993, 185-193.

20. R.A. Botafogo, E. Rivlin, and P. Shneiderman, Structural Analysis of Hypertexts: Identifying Hierarchies and Useful Metrics, ACM Transactions on Information Systems, 10:2, April 1992, 142-180.

21. C. Guinan and A.F. Smeaton, Retrieval from Hypertext Using Dynamically Planned Guided Tours, Proc. ECHT '92, *Association for Computing Machinery*, New York, December 1992, 122-130.

22. J. Allan, Automatic Hypertext Construction, Doctoral Dissertation, Computer Science Department, Cornell University, November 1994.

23. G. Salton and J. Allan, Selective Text Utilization and Text Access, Proc. Hypertext '93, *Association for Computing Machinery*, New York, November 1993, 131-144.

24. G. Salton, J. Allan, C. Buckley, and A. Singhal, Automatic Analysis, Theme Generation and Summarization of Machine-Readable Texts, Science 264, 3 June 1994, 1421-1426.

25. M.A. Hearst and C. Plaunt, Subtopic Structuring for Full-Length Document Access, Proc. SIGIR '93, *Association for Computing Machinery*, NY, 59-68.

26. G. Salton and A. Singhal, Automatic Text Theme Generation and the Analysis of Text Structure, Technical Report TR 94-1439, Computer Science Department, Cornell University, Ithaca, NY, June 1994.

27. G. Salton and J. Allan, Automatic Text Decomposition and Structuring, Proc. RIAO-94, CID, Paris, November 1994, 6-20.

28. D.K. Harman, editor, The First Text Conference (TREC 1) NIST Special Publication 500-207, National Institute of Standards and Technology, Gaithersburg, MD, March 1993.

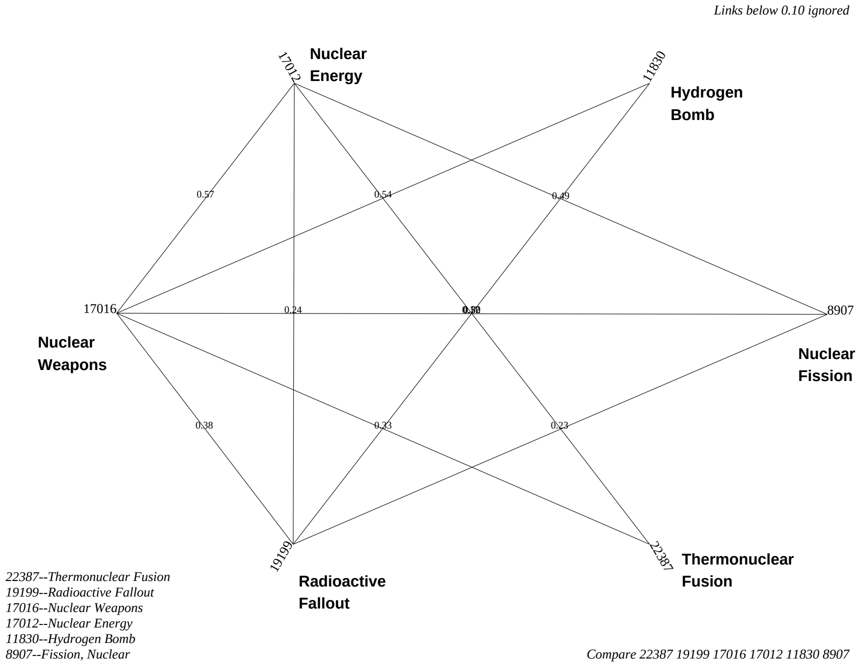

Figure 1. Term Relationship Map

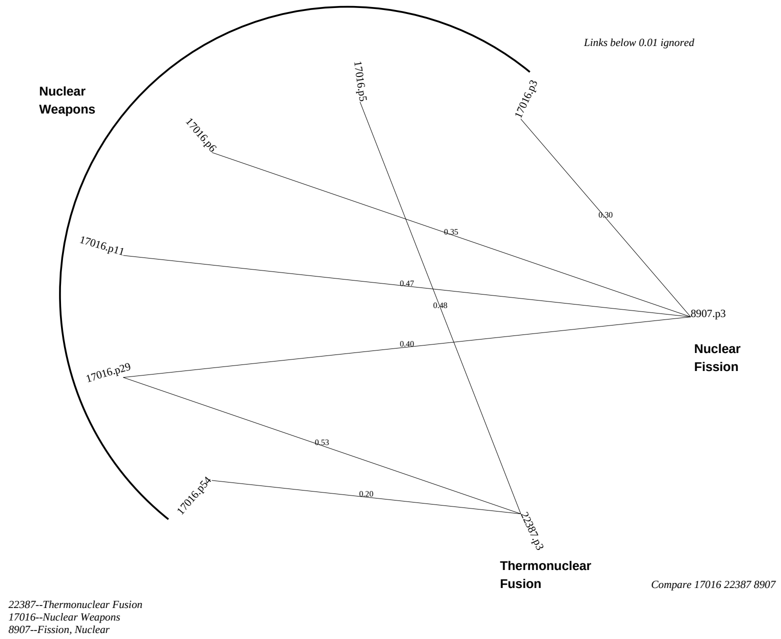

Figure 2. Paragraph Relations

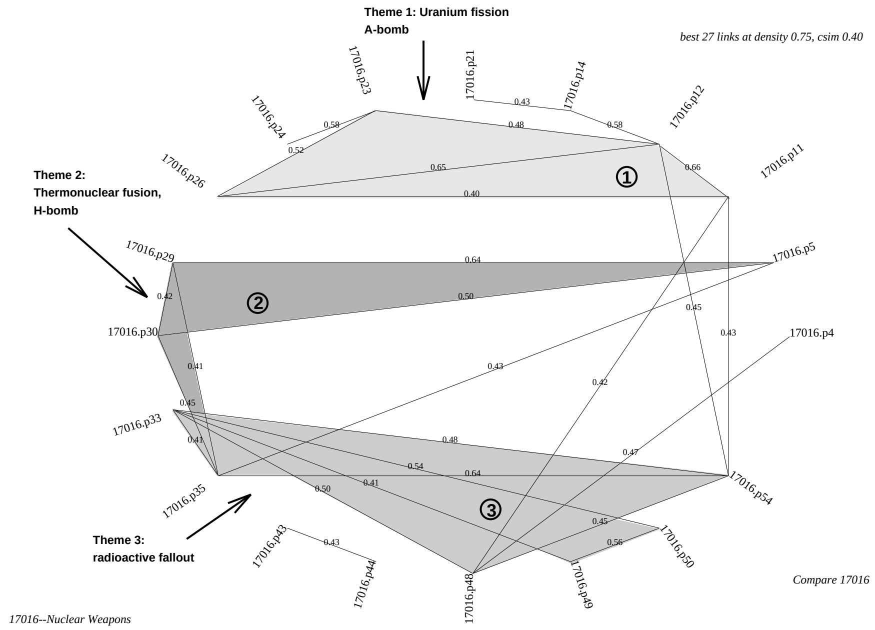

Figure 3. Text Themes (17016 Nuclear Weapons)

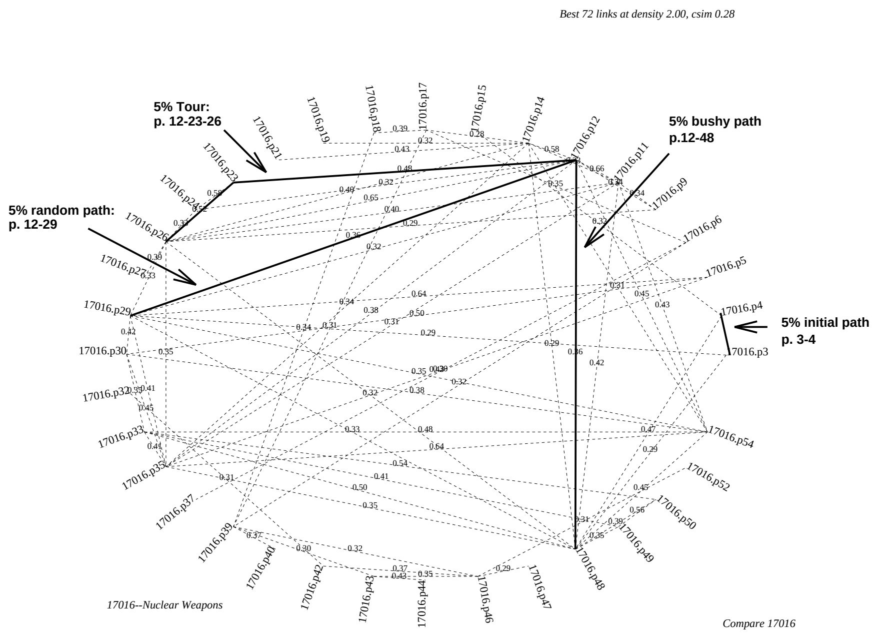

Figure 4. Typical 5% Paths (17016 Nuclear Weapons)

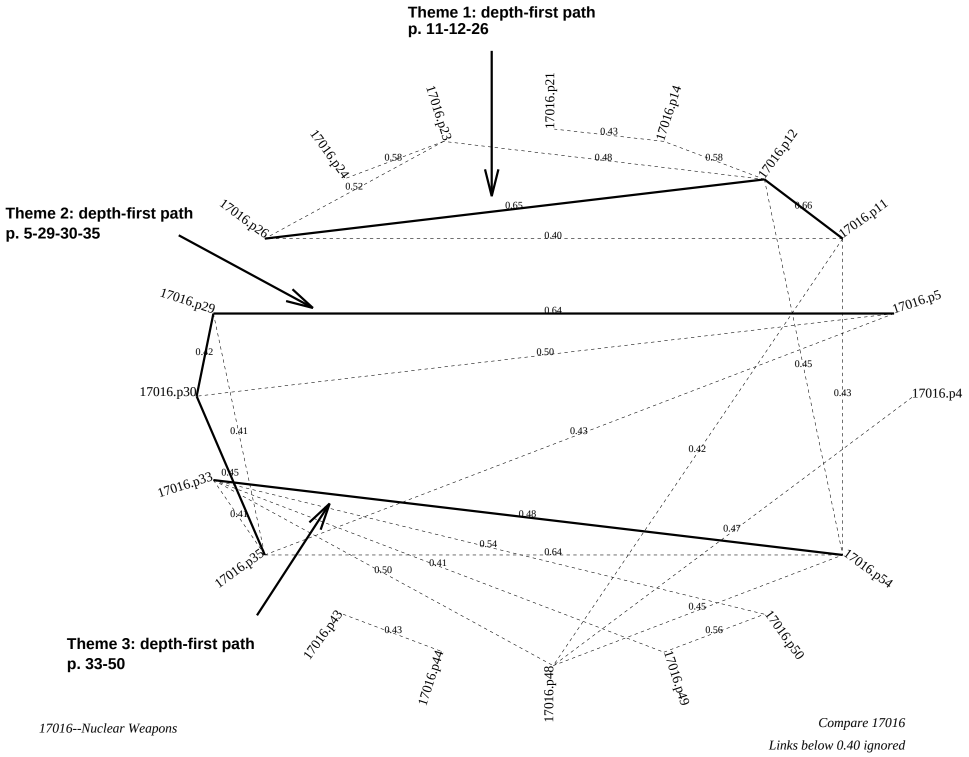

Figure 5. Depth-First Path Within Themes

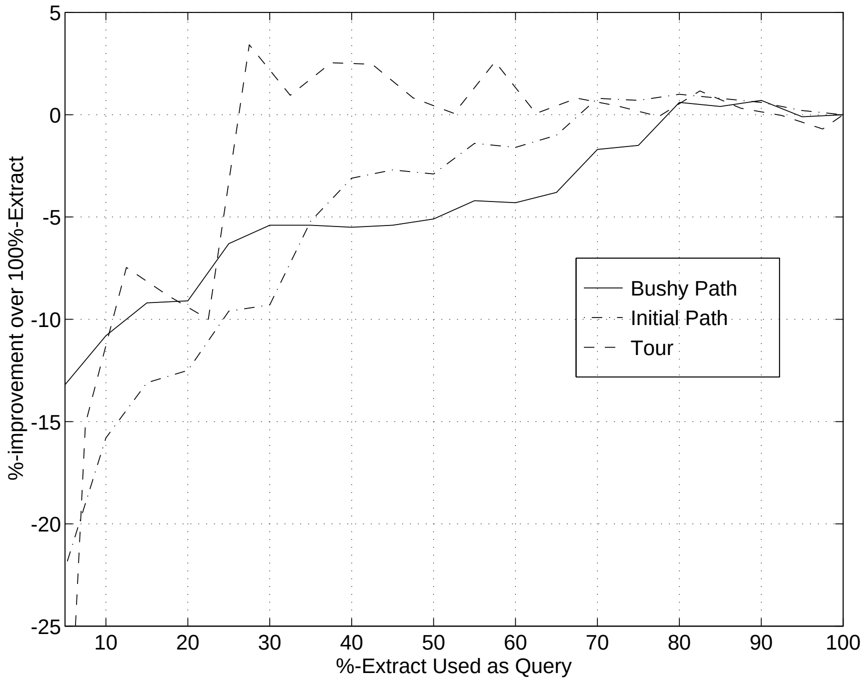

Figure 6. Evaluation of Initial Paths, Bushy Paths, and Tours of Specified Length (comparison with performance of full-length documents averaged for 70 queries and 46,000 documents – Federal Register collection)

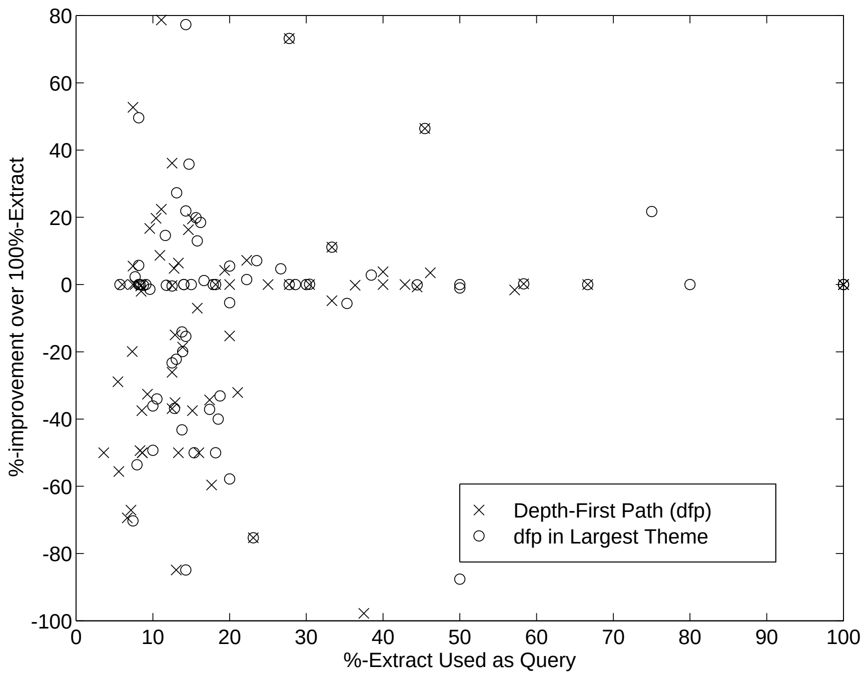

Figure 7. Evaluation of Depth-First Paths and Depth-First Paths Within Largest Theme (comparison with performance of full-length documents averaged for 70 queries and 46,000 documents – Federal Register collection)

1. **Start with a minimal csim threshold (typically 0.01) to insure that all potentially important links are present, and specify a desired map density level (typically 2.0, indicating that the number of allowed links is at most twice the number of nodes).**

2. **Progressively raise the csim value, thus producing sparser maps, until the density restriction (maximum number of allowed links) is satisfied.**

3. **Identify the triangles in the final map and compute a centroid vector for each triangle.**

4. **Merge pairs of triangles if the respective centroid similarity exceeds the tsim threshold (typically tsim 0.30) and continue until no further merging is possible.**

5. **Identify each merged triangle set with a text theme.**

Table 1. Text Theme Generation Process

Type of Excerpt	Construction	Characteristics
x% initial path	first x percent of text units such as paragraphs or sentences	covers beginning of text including abstract, summaries and introductory material
x% bushy path	best x percent of text units in decreasing order of number of links to other text units	covers important text units considered central in the text; excerpts may not represent a coherent whole
x% random path	x percent of text units chosen randomly	generally disconnected units of text
x% tour	x percent of most linked text units included in theme definitions	covers theme information but short tours may not cover all themes
depth-first path	a continuous path starting with first important node and proceeding in depth-first order	represents a coherent text excerpt but may not be comprehensive

Table 2. Characteristics of Important Text Traversal Paths

5% Initial Path p3-4

- **p3** **Nuclear Weapons, explosive devices, designed to release nuclear energy on a large scale ...**
- **p4** **Nuclear explosives involve energy sources within the core, or nucleus, of the atom ...**

5% Bushy Path p12-48

- **p12** **The light isotope of uranium, uranium 235, is easily split by the fission neutrons and, upon fission, emits an average of about 2.5 neutrons.**
- **p48** **The residual nuclear radiation, generally known as fallout, can be a hazard over very large areas ...**

5% Random Path p12-29

- **p12** **The light isotope of uranium is easily split by the fission neutrons ...**
- **p29** **Even before the first atomic bomb was developed, scientists realized that a type of nuclear reaction different from the fission process was theoretically possible.**

5% Tour p12-23-26

- **p12** **The light isotope of uranium is easily split by the fission neutrons ...**
- **p23** **The fissile uranium 235 isotope accounts for only 0.7 percent of natural ura nium; the remainder is composed of the heavier uranium 238.**
- **p26** **Although the heavy uranium isotope uranium 238 will not sustain a chain reac tion it can be converted into a fissile material by bombarding it with neutrons ...**

Table 3. 5% Text Excerpts (17016 Nuclear Weapons)

Theme 1: Depth-First Path p11-12-26

- **p11** **When the uranium nucleus fissions, it breaks up into nuclear fragments and releases energy ...**
- **p12** **The light isotope of uranium is easily split by the fission neutrons ...**
- **p26** **Although the heavy uranium isotope will not sustain a chain reaction, it can be converted into a fissile material ...**

Theme 2: Depth-First Path p5-29-30-35

- **p5** **Other types of bombs were developed to tap the energy of light elements such as hydrogen ...**
- **p29** **Even before the first atomic bomb was developed, scientists realized that a (fusion) type of nuclear reaction was possible ...**
- **p30** **However, development of the hydrogen bomb was impossible before the perfecti on of A-bombs.**
- **p35** **The thermonuclear bomb that exploded in 1954 was a three-stage weapon ...**

Theme 3: Depth-First Path p33-50

- **p33** **The March 1954 explosion led to worldwide recognition of the nature of radio active fallout ...**
- **p50** **Human experience with radioactive fallout has been minimal. Principal known case histories are confined to accidental exposure of natives and fishermen**

Table 4. Excerpts for Depth-First Paths Within Themes

Type of Path	Short Path	Long Path
Initial Path	coherent non comprehensive includes unimportant material	coherent comprehensive includes unimportant material
Bushy Path	not coherent not comprehensive important excerpts only	not coherent comprehensive important excerpts only
Tours	may be coherent not comprehensive confined to one topic important excerpts	may be coherent comprehensive several topics important excerpts
Depth-First Path	coherent not comprehensive single topic in general	coherent not comprehensive often single topic only
Paths in Themes	generally confined to one topic	generally confined to one topic

Table 5. Path Characteristics