Using Librarian Techniques in Automatic Text Summarization for Information Retrieval

Min-Yen Kan
Department of Computer Science
Columbia University
New York, New York 10027
min@cs.columbia.edu

Judith L. Klavans
Center for Research on Information Access
Columbia University
New York, New York 10027
klavans@cs.columbia.edu

ABSTRACT

A current application of automatic text summarization is to provide an overview of relevant documents coming from an information retrieval (IR) system. This paper examines how CENTRIFUSER, one such summarization system, was designed with respect to methods used in the library community. We have reviewed these librarian expert techniques to assist information seekers and codified them into eight distinct strategies. We detail how we have operationalized six of these strategies in Centrifuser by computing an informative extract, indicative differences between documents, as well as navigational links to narrow or broaden a user's query. We conclude the paper with results from a preliminary evaluation.

Categories and Subject Descriptors

H.3.7 [**Information Storage and Retrieval**]: Digital Libraries—*User Issues*; I.2.7 [**Artificial Intelligence**]: Natural Language Processing—*Text Analysis*

General Terms

Algorithms, Design, Human Factors

Keywords

Reference Librarian Techniques, Automatic Text Summarization, Information Retrieval User Interfaces

1. INTRODUCTION

With the increasing amounts of resources being indexed and provided in the digital library, search techniques need to enable users to overview portions of the collections as well as pinpoint particular resources to fit their information needs. Traditional paper libraries have had these problems since their conception. In this paper, we examine and categorize eight strategies that reference librarians use in assisting information seekers both in-person (during informational reference interviews) and offline (by authoring subject guides).

In a digital library context, search is done electronically, which alters the way search is conducted. We can still use many of same techniques as in the paper library but need to adapt them to the online context. We discuss Centrifuser, a system that we developed that operationalizes six of these librarian strategies. Centrifuser receives a list of relevant documents from a digital library or standard information retrieval (IR) engine as input, and uses automatic text summarization (ATS) techniques to post-process the documents into a summary form. Centrifuser uses ATS to provide an overview of documents returned by a search query that is constructed by extracting prominent sentences, as well as indicative summaries that differentiate the documents from each other that are constructed by natural language generation techniques.

2. MATCHING THE RETRIEVAL TASK WITH ASPECTS OF SUMMARIZATION

It is well understood that people use documents for different purposes. It is likewise the case that there should be different types of summaries to support these purposes. We have examined three specific dimensions of ATS that are of particular importance for the purpose of summarizing documents for online search and retrieval:

1. **Favoring multidocument over single document summarization**: Most ATS systems function in the single document context, where a single document is condensed to a shorter form. For a good overview of single document summarization technology, see [23]. In the context of information retrieval, we have multiple documents that are returned by a single search request. To generate a single output that summarizes the salient points across these multiple documents is more difficult. Since the documents are related by a common query, they likely contain similar content; thus a system cannot simply concatenate many single document summaries together, because repetition of salient points would result. If ATS is to be a successful methodology for information retrieval, a system that can handle repetition in multiple documents is a prerequisite. CENTRIFUSER's components examine similarity and differences among each document's

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for pro@t or commercial advantage and that copies bear this notice and the full citation on the @rst page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior speci@c permission and/or a fee.
JCDL'02, July 13-17, 2002, Portland, Oregon, USA.
Copyright 2002 ACM 1-58113-513-0/02/0007 ...$5.00.

structure, so it is designed specifically to handle multiple documents.

2. **Favoring query-based over generic summarization**: ATS systems often produce generic summaries that highlight the most salient points of a given text. However, in the online search and retrieval context, an ATS system has access to the query given by the user and should adapt its output to suit the user's declared information need. There are many instances when a rational IR framework finds query keywords in only a subsection of a larger document. Showing this relationship between the query terms and the document has been proven to be an important factor ([1] shows the difference between AltaVista – which does generic summarization by reporting the n-top sentences of a document – and Google – which reports sentences in Query Word In Context format). It has been a focus of research in graphical user interface design [10]; query-based text summarization is the logical parallel in ATS. While it would be acceptable to store generic document summaries and present them in an IR system, a more favorable approach is to produce per-query customized summaries. This latter approach is the method used by CENTRIFUSER.

3. **Informative versus indicative summarization**: Informative summaries provide information on the salient aspects of a document, seeking to cover as many topics as possible. These summaries omit detail or supporting information and just cover the most important points of the document. Summaries of this type often are used in place of the document as an overview, and are suitable for fulfilling a user's information need if they are browsing for information or have a general interest in the subject of the document. An example of this is an article abstract that mentions the problem, methodology and results of the article.

   Indicative summaries, on the other hand, are meant to only hint at the contents of the document. In the IR context, indicative summaries play an interesting role because they help the user in judging the relevance of the document, and in determining whether to consider full-text retrieval. They assist a user who is searching for information and has a specific information need. We can think of the indicative summary as a vehicle for routing a user to a specific document in the query result set.

   The type of summary that should be produced by an ATS system for IR depends on the scope of the user's information need. Indicative summaries are important to searchers, while informative summaries are more important to browsers. Both modes of information access will occur in the digital library, so unless the user explicitly gives a preference, both modes should be utilized. An ATS system for IR can either match together two separate systems that construct indicative and informative summaries, or try to implement a single system that produces both. CENTRIFUSER produces both indicative summaries for searchers and informative summaries for browsers using the same underlying framework.

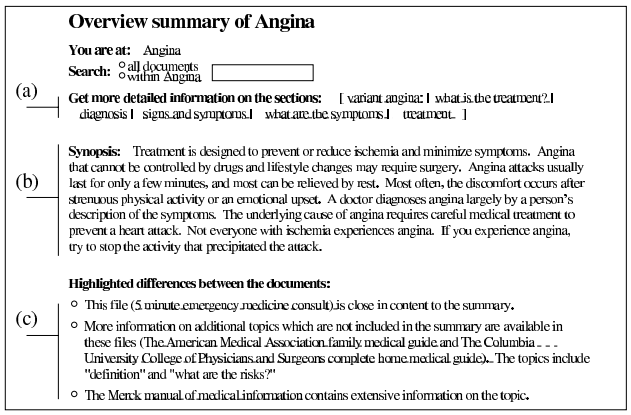

Figure 1: A CENTRIFUSER summary generated from post-summarizing the first ten hits from Google on the query "Angina".

3. CENTRIFUSER - ATS USING SIMILARITIES AND DIFFERENCES

CENTRIFUSER (loosely short for *centroid fuser*) is an implemented system that we developed using automatic summarization techniques. CENTRIFUSER was developed within the scope of the PERSIVAL medical digital library project (DLI-2) at Columbia University [19]. CENTRIFUSER acts as a post-processing system for a standard framework information retrieval system (i.e., reports results in a ranked list format) which takes the first ten or twenty most relevant search results and produces summaries.

Currently, CENTRIFUSER provides three different types of information as output, shown in Figure 1. At the top of the output are (a) **navigation links** that allow the user to navigate to broader and narrower subjects related to the original query. In the middle, (b) an **informative extract** organizes sentences from the search results that represent their similarities. Finally, the bottom contains (c) **indicative group summaries** which collect the search result documents into several text bullets. The texts for the bullets are completely computer generated and based on differences in the document's topical structure and distribution.

CENTRIFUSER creates these multidocument summaries by utilizing the basic strategy of using similarities and differences between the documents [18]. The same paradigm has been used for the visualization of retrieval results in graphical user interfaces (e.g., BEAD [6]). CENTRIFUSER differs from previous research in multidocument text summarization in that the system additionally compares documents against a model of expected information for documents. By this, we mean that CENTRIFUSER learns information about typical document length, organization and subtopics for particular types of documents, and codes this knowledge in a resource called the Composite Topic Tree (CTT). A composite topic tree that we derived for patient information documents on diseases shows that symptoms, diagnosis, treatment, and links for more information are typical subtopics and usually occur in the specified order is shown in Figure 2. The CTT additionally contains subtopic frequency information that enables the system to detect common or rare subtopics in documents, which can be conveyed to the user.

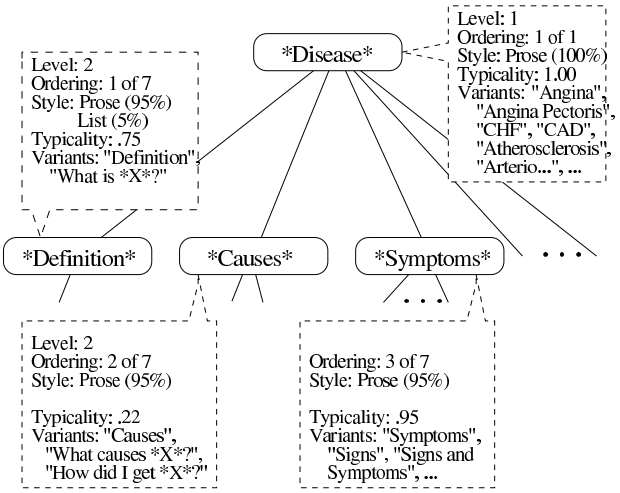

Figure 2: A sample composite topic tree for patient information on diseases.

The actual implementation of the system is described in [12]. We overview the steps that CENTRIFUSER takes in constructing the summaries.

1. A set of topically related documents is selected by a search engine. These documents constitute the input to CENTRIFUSER.

2. Each document in the set is analyzed in order to identify its structure, in terms of subtopics. These analyses form the basis for transformation of each document into a hierarchical tree of topics. For example, a document containing patient information for a particular disease will consist of subtopics including symptoms, diagnosis and treatment. In a related process outside of this summarization overview, CTTs are constructed by merging together many of these document topic trees.

3. A user query is matched to a particular subtopic in each document. This match divides the document topic structure into three different subtopic regions to use for tailored summarization: the relevant region, the irrelevant (usually broader) region, and the intricate (too narrow) region. This is used as the basis for comparison across documents. Figure 3 illustrates this process.

4. The relevant regions are processed in two ways. First, relevant regions are linked across documents by subtopic. For example, all symptom topics are linked. The topics are clustered [9] and similar propositions are identified. A representative sentence from each topic is collected and serves as the informative extract, which is shown in Figure 1, as item (b). Second, relevant region subtopics are classified by statistical methods to determine importance, some resulting in the most common or typical nodes, and the others being classified as rare or atypical.

5. Ratios of the number of typical, rare, intricate and irrelevant topics are computed and used to determine a document category. These categories reflect the topical distribution of each document. For example, a document might focus on symptoms, with little information on prognosis.

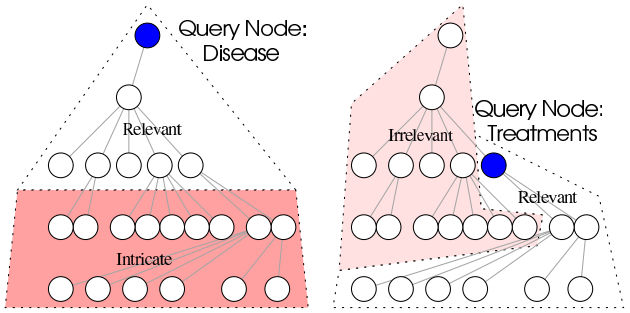

Figure 3: The three subtopic regions as defined by the query.

6. For each document category, an indicative summary is generated, shown as item (c) in Figure 1. These indicative summaries also include metadata about different types of content (e.g. if the document contains figures or tables).

Thus, CENTRIFUSER uses the topical structure of related documents to compute a summary. These summaries are tailored to the user query. For example, if the user wants to know about symptoms, then a summary which contains more information on symptoms will be generated.

4. USER INTERFACE DESIGN BASED ON LIBRARIAN TECHNIQUES

The design of CENTRIFUSER's output draws on theory from two different areas of information science. First, we reviewed theory on the informational reference interview process, which led us to enumerate how strategies from the interview process can be adapted to the IR context. This new view of this existing information has permitted us to design CENTRIFUSER in a motivated way, taking into account user needs as part of the system design process. Second, we examined cataloguing resources, such as card catalog entries and subject guides. A review of the prescriptive guidelines for writing these resources along with a corpus analysis of actual cataloguing records also yielded strategies that reflect the cataloger's knowledge. These strategies were adapted in the implementation of CENTRIFUSER. Thus, we have incorporated research from two key areas of information science: informational reference interview and cataloguing resources.

5. INFORMATIONAL REFERENCE INTERVIEWS

It is a well known fact that users often come to the reference desk in search of many different things: directions to the restroom (directional queries), instructions on photocopier use (instructional queries), requests to borrow books (circulation queries), and assistance with finding information (informational queries) [13]. This final category of dialogue, called the *informational reference interview*, is a process that can be factored into four steps:

1. Communicating the information need;
2. Selecting a search strategy;
3. Executing the search;
4. Assessing the relevance of the results.

Katz [13] begins his discussion on this subject by stating that "The most common complaint heard among reference librarians about their work is that few people know how to ask reference questions." Thus, the first and possibly most difficult part of the informational reference interview is to find out what information the user wants.

A primary difficulty is that users often underspecify their information needs. An example of this is when a user comes to the reference desk with a very specific information need (e.g. recent years' statistics on number of people in America afflicted with high blood pressure) but only communicates a part of it to the librarian in their query (e.g. "I need information on high blood pressure"). When this occurs, and unexpected results are returned, we find that a) the user made a tacit assumption that the librarian understood the specifics implied by the request, or because b) the user lacked the terminology to correctly express their information need.

Strategies for addressing this issue can also be separated along this distinction. When the librarian feels that the query is underspecified because its purpose or scope isn't clear, she may want to explicitly ask for the user's intentions ("Are you looking for information on high blood pressure for a report?"). A user may be sensitive about his needs (e.g. if the information is for a family member who is ill) and thus questions of this sort need to be phrased in a manner that make it clear that a response is optional and that the purpose of the question is to target relevant sources [11]. When the librarian feels that the query is underspecified because the user does not possess the knowledge to properly express their needs, she can enumerate possible subtopics or facets of the user's query ("Well, there's information about hypertension on a number of subtopics, including symptoms and treatments."). This cooperative type of response can aid the user in better defining his query for the librarian to understand.

Once the needs of the user have been successfully communicated, the librarian's task is to translate the information need into a search plan. What information sources should be consulted? Should a general catalog search be done or should specific reference materials for a particular subject area be consulted? Which materials may yield higher quality results? Which searches are easiest and least costly to perform? Based on the knowledge of the scope and type of information needed by the user, the librarian can assist in formulating an appropriate search strategy.

In the course of, or at the end of the search process, the librarian may need to evaluate the materials produced by the search. Often, these search results are better assessed by the librarian than the user. Librarians are often able to quickly match the genre of information needed to the specific user type. Does the user want information found in newspaper clippings, trade magazines or scholarly journals? Is the source an authority on the topic and is it highly regarded by scholars in the field? Is the source an up-to-date resource for the topic? In this regard, the librarian often acts as an information filter, narrowing choice down to a manageable set of alternatives (2-5 sources). Occasionally, librarians are called on to recommend their favorite source and in these cases a single best choice is picked from all the alternatives.

5.1 Reference Interview Strategies

To summarize, there are many aspects of the reference interview that exhibit specific strategies that would be desirable to adapt to a digital library setting. These include the following (non-exhaustive) features that we have discussed:

1. Reference librarians may first need to clarify and elicit the user's information needs;
2. Reference librarian may need to decide on which information sources to consult based on an assessment of the user's need;
3. Reference librarians must be able to evaluate potential sources during the formulation of the search process and during retrieval;
4. During an interview, reference librarians also form and take advantage of a user model and his information need.

Let's examine how CENTRIFUSER supports each of these strategies:

5.1.1 *Clarify underspeci®ed queries*

CENTRIFUSER attacks the problem of underspecified queries by generating navigation links. As seen in the example, a simple query on the health condition of "Angina" generates navigation links that assist the user in verifying that his query was correctly understood by offering related and narrower subtopics. These links also help in making the user's information need explicit: if the user was looking for information on "treatments for angina", "treatments" now appears on the navigation controls, which the user can click on to run the specified follow-up query.

5.1.2 *Form a search strategy*

During the reference interview, the librarian formulates a search procedure based on the criteria the user provides. This search procedure may touch upon resources physically located in different places or which differ in methods of access. In the online context, typical IR systems consult a single (possibly huge) database, thus alleviating the need to use this strategy. Metasearching systems have a parallel problem, in that they need to evaluate multiple resource access and access cost [7]. However, as CENTRIFUSER is designed as a post-processor for a standard IR engine, it is not actually involved in the search process, so this strategy is not directly applicable.

5.1.3 *Evaluate documents for initial content relevance*

CENTRIFUSER's indicative group differences also fulfill this strategy. The grouping of documents is done by topical distribution, such that documents that list common topics (i.e., topics that are known to be frequent and thus important) are gathered together into a single bullet. Similarly, documents that contain mostly rare (i.e. infrequent) information or which best address narrower or broader topics are also grouped together. By bringing out the salient differences

between documents, the user gains a better perspective of the subtopics available in the document groups and can best make a relevance assessment on a per group basis. This speeds up the evaluation process, since users only need to evaluate each grouping for relevance, rather than evaluating each document individually. Figure 4 shows the anatomy of a generated (exploded) individual text bullet in detail.

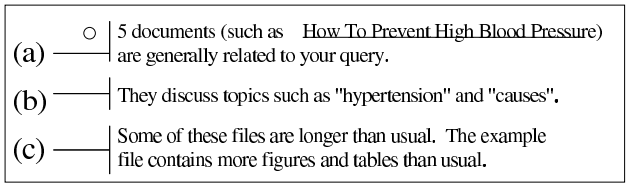

Figure 4: A single bulleted indicative summary generated by CENTRIFUSER, consisting of (a) a category description, (b) treated subtopics, and (c) metadata features that differ from expected.

5.1.4 Form a user model

CENTRIFUSER addresses an aspect of the user model that is different from traditional views. The summaries tailor information for two different modes that users employ for information access: browsing and searching. The informative extract provides an overview of similarities across the documents, and is constructed by extracting sentences from the most frequent (i.e., important) subtopics of the documents. The system uses a concept clustering technique to group together sentences with similar information. By extracting one sentence per cluster, we avoid the problem with repetition of information. As stated in Section 2, an informative summary best supports the information access mode of browsing for information at a high-level. The indicative group summaries serve the complementary information access mode of searching.

5.2 Adapting strategies to the IR context

It is important to note that these interview strategies cannot be brought into the digital library wholesale. They must be adapted to operate within the confines of IR. Reference interviews take place in the physical world, where speech and other non-verbal cues aid the librarian in constructing a user model and in understanding and presenting feedback to the user. In the IR context, users are limited to using keyboard and mouse for input, which exacerbates the communication problem. Since input is so difficult to provide, users often underspecify their information need. Thus the system must be able to provide useful feedback to the assist the user in assessing relevance and clarifying their queries. In an online search, users have also come to expect near-instantaneous feedback during the entire search process. Thus, search requests are often iterative because there is little time cost; initial searches often probe for terms to use in follow-up searches. This is in contrast to the distinct, segmented steps of the reference interview in Section 5. The implication of this is that at all times during the course of a search in the digital library, the user interface must support all of subtasks simultaneously: including information need clarification, feedback, and evaluation of results.

6. CATALOGUE INFORMATION

Cataloguers also assist users in finding information in the library, in a complementary manner. Cataloguers compile card catalog entries and subject guides, which are a written form of the information search and retrieval process. They differ from their informational reference interview counterpart in that they are static and cannot participate in an interactive dialogue with the user; the text has been written even before the user first encounters the resource.

This asymmetry has several ramifications for subject guides devoted to assisting search: a) resource descriptions should contain appropriate information to perform a relevance judgment; b) differences between resources should be made clear; c) the organization of the entries must be geared towards a specific method of search; d) methods of locating entries by alternative search methods or orthogonal classification axes should be provided. We now examine each of these areas of concern in more detail.

a) **Descriptions help to perform relevance judgment.** The *Guide to Reference Books* [3], perhaps the reference librarians' most valuable tome, states that its annotations "enable users of the *Guide* to understand a source's scope and visualize its arrangement and features so that it can be used more easily at the shelf. In writing annotations, contributors were asked to consider a [book]'s purpose, audience, scope, coverage, arrangement and special features, and were invited to comment on its useful in reference." This description points out that meta-information such as purpose, audience, and coverage are important features in providing a description for the user (even when the user is a librarian herself), in addition to direct information on source content.

b) **Differences between resources should be made clear.** Many subject guides list several different resources for an area. Multiple resources can be listed for completeness, but are often listed because the different resources are authoritative for separate subtopics (e.g. "*A* contains more information about treatment options, but consider using *B* if you are trying to find information about procedures for diagnosis") or better suited for certain purposes (e.g. "*A* is better for its elegant pictures, but *B* contains a much better narrative").

c) **Organization assists a specific method of search.** Different subject guides that address different areas are often organized different to best serve their consumer population. An interesting example of this is the *Art History Resources on the Web* subject guide [25], which organizes art history resources at its top level in chronological order for Western art, but geographically for most other regions. The organization shows that even within a particular field there can be orthogonal ways of structuring information that are equally valid.

d) **Support alternative search methods or orthogonal classification axes.** Large subject guides – such as the aforementioned *Guide* – in addition to being logically organized, also often have an index to support keyword search. Some resources provide multiple indices on their information content (e.g. Chinese dictionaries that allow indexed search by radical, pronunciation and number of strokes) that enable multiple points of access to the same information.

In addition to identifying these four strategies as motifs in the compilation of subject guides, we also examined the catalog entries. We did this along two different dimensions: a) single versus multiple document summaries; and b) ex-

amining prescriptive guidelines (from a top-down perspective) versus descriptive corpus analysis (from a bottom-up perspective). We will discuss this work first from the single document perspective, then from the multiple document perspective.

6.1 Single document analysis

Cataloguing records are probably the best example of indicative summaries in the context of the physical library. We can think of these cataloguing records as indicative summaries that assist the user in making a relevance judgment about whether a book or resource is useful enough to go to the library shelf to locate it.

We examined what types of information should be present in a single document indicative summary by looking at guidelines for writing these card catalog summaries. Both the MARC cataloguing guidelines [15] and the ANSI 1979 [2] guideline for book jackets contain such information. For example, information on the topic of the book should be included in a summary. This can be presented in a number of ways: by listing the salient topics in the table of contents of the book, by presenting keywords, or by writing a summary of the book. Other types of information also can play a role in assisting the user in making a relevance judgment and also are sometimes included in card catalog entries. This includes information such as the author, the book's edition, its purpose, and its targeted audience.

Based on the guidelines, we subdivided the list of *document features* into two broad categories of information: a) document-derivable information, and b) metadata information. Information that can be deduced from the text of the document itself belong in the first category. This includes information about the topic and structure of the document (which our system currently handles via the construction of the document topic tree). Metadata information includes everything else: information that might not be present or deriveable from the text of the document. These metadata document features include information such as the document's title, its authors or editors, as well as purpose or audience information.

Knowing what types of information to place in a summary solves a part of the problem. However, we do not have an idea of which of these document features are more important than others. To address this problem, we used a more detailed corpus analysis.

The first part of this task was to locate and collect a corpus. We used Columbia's local catalog, CLIO (Columbia Libraries Information Online), to locate indicative summaries. CLIO uses the Library of Congress' MARC cataloguing interchange format. MARC records consist of many structured fields, including the *520 field*, which is an optional field that contains a text summary. In our local catalog, the MARC 520 field is shown to the user when one is provided in the catalog data. It should be noted that many of the 520 fields in the Columbia catalog are typed-in book jacket texts, and may have been purchased from an outside vendor. Our point here is that the MARC 520 summaries are often independently constructed, and their format and content do not take into account the presence or absence of other MARC fields. For instance, even though the title of a book is included in an online catalog view of a book, the MARC 520 summary field may include it anyway, because it probably was constructed independently of the online catalog display.

Since we are working within the medical domain, we ran several medical queries through our library catalog to generate a large set of relevant records. Of these records, 82 of these had the MARC 520 field, and were extracted to build an initial corpus for the analysis.

Our task was to find the importance of each of the document features expounded in the MARC and ANSI guidelines. Finding this information would allow us to replicate the same information flow in our own automated indicative summarization program. We first manually assigned each sentence in each of the 82 summaries to one of the document feature categories. Sentences that did not fit well in any feature were placed in an "other" category, while sentences that addressed multiple features were assigned to each individual category. As multiple featured sentences were placed in multiple categories, this was a "soft" classification task. We then calculated the frequency of each document feature. This information is shown in Table 1.

Document Feature	% appearance in corpus
Document-derived features	
Topicality (e.g. "topics include symptoms, ...")	100%
Content Types (e.g. "figures and tables")	37%
Internal Structure (e.g. "is organized into three parts")	17%
Readability (e.g. "in plain English")	18%
Special Content (e.g. "Offers 12 credit hours")	7%
Conclusions	3%
Metadata features	
Title	32%
Revised/Edition	28%
Author/Editor	21%
Purpose	18%
Audience	17%
Background/Lead	11%
Source (e.g. "based on a report")	8%
Media Type (e.g. "Spans 2 CDROMs")	5%

Table 1: Distribution of document features in library catalog summaries of consumer healthcare publications.

We drew two analytical conclusions based on the data. First, all indicative summaries contain some indication of topic or content. This is not surprising, but suggests that summaries should be organized around this important (i.e., common) feature. Second, metadata features contribute significantly to indicative summaries. In fact, if the metadata features were independently distributed, each indicative summary would contain at least one metadata feature. This tells us that we cannot rely purely on extractive techniques to build indicative summaries. We must utilize metadata present in library catalog resources.

6.2 Multiple document analysis

To our knowledge, there is no publically available corpus of indicative multidocument summaries, so we could not perform a descriptive study as we did with single document summaries. However, we did find prescriptive guidelines about how people should write multiple single-document in-

dicative summaries that relate to a single topic. The Open Directory Project (ODP) [22], is a website hierarchy project that is similar in nature to Yahoo!. The ODP is a browsable topic hierarchy that organizes websites that address a particular topic into a single category page, which are maintained by human editors. In the ODP editor's guidelines (which are required reading for all editors) they state that in building each category page, editors should "make clear what makes a site different from the rest". This tells us that differences are very important to multiple document summaries, especially if the documents that are input for the summary are on the same topic.

6.3 Study Implications to Centrifuser

This initial study of the indicative summaries helped us formulate the design of the indicative summary module in CENTRIFUSER. Because of overwhelming importance of the topicality document feature in the single document corpus analysis, we use subtopic structure and distribution as our primary means of distinguishing documents from each other. Furthermore, as metadata features play an important role, we include these features in our indicative summaries when space allows and when these particular document features differ from expected (when the value for the particular document(s) differ from the value in the Composite Topic Tree). This last restriction is important to certain types of documents: for example, knowing that a medical book has images may be important (since they mostly don't); also knowing that a book about maps does not contain images is also important (since they mostly do).

Let us now examine how these design decision address the issues that were brought up in the analysis of subject guides, mentioned at the beginning of this section:

6.3.1 *Descriptions help to perform relevance judgment*

CENTRIFUSER's indicative grouping of differences between the documents assists the user in deciding whether a document group fits his information need. The system's algorithm in creating the text for these bullets also includes the addition of metadata features (such as whether the source is particularly lengthy, or possesses images or tables) when the metadata feature varies from the expected value of the feature (see item (b) in Figure 4). This is calculated by comparing the feature for the individual documents against its typical value that is stored in the composite topic tree resource. Recall that metadata is extensively used in construction of the cataloguer's offline subject guide.

6.3.2 *Differences between resources are explained*

The system's indicative group summaries again assist with this concern. Each category that has been assigned one or more documents will be realized with a textual bullet. The bullet includes an explanation of the individual category (item (a) in the above Figure 4) as well as embedding a link to one or more documents that belong to the category.

6.3.3 *Organization assists a speci®c method of search*

As stated earlier, the indicative group summaries are created primarily to support the information access mode of searching. The individual bullets in the indicative summaries are ordered by their relative similarity to the expected information, as calculated by the model of expected information, the composite topic tree. Document categories that are closer to the norm are ordered to the top of the list. Document categories that express rare information or information outside the scope of the query are relegated to the end of the list. Table 2 gives the general description for the categories currently used by the system in the linear order that they are presented in the summary (again, for specifics on classification, see [12]).

Document Category	Dominant Characteristic
1. Prototypical	The typical document, which is well-represented by the informative extract
2. Comprehesive	More than just the typical topics
3. Specialized	Some of the typical topics
4. Atypical	Contains rare information
5. Deep	Contains content that is too detailed for this query but good for follow-up, narrower queries
6. Irrelevant	Mostly contains information outside current query focus but good for follow-up, broader queries
7. Generic	No strong trends

Table 2: Conditions used to categorize documents into document types.

6.3.4 *Support alternative search methods or orthogonal classi®cation axes*

CENTRIFUSER classifies documents by their topic distribution to generate the indicative group summaries, but does not currently allow for alternate means of grouping. Future work would allow the resorting of the documents by alternative criteria (e.g. metadata features: "which documents have more figures or less figures than expected?"), which would in turn generate a rewriting of the text bullets.

Table 3 summarizes our findings. Generally speaking, informational reference interviews exhibit the most flexibility in modalities of input and output, and cataloguing materials, such as subject guides and annotated bibliographies display the least (since they are static).

7. RELATED WORK

In this paper, we have considered information retrieval as a process of information gathering done by people from a cognitive engineering perspective [24], and have presented automatic summarization technology as a method that can incorporate library science techniques.

A complementary approach is to observe library users as they use existing electronic retrieval systems. By interpreting users' behavioral patterns as they employed retrieval systems, more effective user interfaces can be developed to aid the user in using the retrieval system's features. This avenue has been taken by a number of studies [4, 5, 21], which have shown how different aspects of the rational framework can be modified to enhance retrieval effectiveness. Koenemann and Belkin's work [14] is an instance of this, in which the existing rational framework feature of relevance feedback was studied in different forms and shown to positively influence retrieval effectiveness when users understood the technique and controlled its parameters.

We have also examined how reference librarians and cataloguers provide different types of access points to books

Reference Librarian Strategy	Changes necessary with respect to an online IR context
Informational Reference Interview	
1. Clarify underspecified queries	Restricted modality exacerbates this problem
2. Route query to appropriate sources	Not needed in single source IR
3. Evaluate documents for initial content relevance	System should continually support this
4. Form a user model of user	Restricted modality exacerbates this problem
Subject Guides and Annotated Bibliographies	
5. Descriptions help to perform relevance judgment	Should be dynamically calculated
6. Differences between resources are explained	Should be dynamically calculated
7. Organization assists a specific method of search	Should be dynamically calculated
8. Support alternative search methods or orthogonal classification axes	Should be dynamically calculated

Table 3: Librarian search strategies and restrictions from the online information retrieval context.

and other resources. Several studies have focused on the particular access point of written abstracts for expository papers. Fidel [8] examined abstracting policies used by different providers of bibliographic material, and found that many use controlled vocabulary for terminology and guidelines were followed for including metadata document features. Liddy [16] explored how abstracts' discourse structure could be detected and exploited in the context of online IR.

Much work has also been done on graphical user interfaces for IR. Text-centric user interfaces (such as ones produced by Centrifuser) are advantageous in that they can be readily understood by lay users. Additionally, text can be customized to serve different purposes: to offer overview summaries for browsers and to highlight differences between documents for searchers. Recent work [20] compares textual, iconic, graphical and "spring"-type displays in effectiveness, in which icons and text were found most effective. Although Morse and colleagues found that their text interface was disliked, we believe this may be due to the format of the text presented, not the medium of text itself.

8. EVALUATION

The motivation for developing CENTRIFUSER was to study ways to provide patients with access to consumer health information. As such, the initial evaluation of CENTRIFUSER also uses patient health information documents. In our first round of evaluation of CENTRIFUSER, we have followed a think-aloud protocol by asking patients' family members *in situ* at the hospital setting to comment on CENTRIFUSER and three other user interfaces. We selected these three other systems based on differences in levels of automation. They are: 1) About.com, a human expert site, in which human experts have built web sites for particular topic categories and provided original content and links to other sites; 2) Yahoo!, a web site hierarchy that uses human editors to approve new site descriptions and links; and 3) Google, a fully automated search engine. Since CENTRIFUSER is a post-processor for a standard IR search engine, the CENTRIFUSER system was set up to post-process the Google documents (i.e., CENTRIFUSER's document set was the same as Google's). Identifying images, font and color schemes were omitted or standardized between the systems so that the identity of each system was masked.

As the evaluation utilizes patient interviews and thinking-aloud, our goal was to elicit feedback from the patient's families on areas of strength and weakness. A total of thirteen groups of participants were interviewed and asked to comment about the user interfaces as they were presented in random order. A final ranking survey explicitly comparing the systems on a 7-point scale was also administered at the end of the survey, which asked participants to consider each system's a) content, b) types of information available, c) ease of locating information, d) clarity of follow-up steps, e) layout and f) overall general satisfaction.

Results from the think-aloud sessions have been transcribed but not yet fully analyzed, so we present preliminary results based solely on the ranking survey. Since the main objective of the evaluation was centered on the think-aloud data, we will do extensive analysis on each participant. Figure 5 shows the results of the ranking survey from 11 of the 13 user groups. We had to disqualify two groups from the results because they did not understand the CENTRIFUSER user interface. Our preliminary conclusions indicate that:

1. Human experts perform best. About.com scores above the other interfaces across the board. Thus, human experts can be thought to represent a gold standard for other systems to aspire to.

2. CENTRIFUSER either outperforms or performs equally well as Google. We believe that this finding validates that participants found the post-processing of the same Google results to be more usable in the automatic text summarized form, despite errors generated in the production of the summaries.

3. CENTRIFUSER performs around the same level as Yahoo!. This is an exciting finding, as a fully automated system such as CENTRIFUSER may be able to provide the same level of satisfaction as a resource that expends much human effort.

9. FUTURE WORK

Our future work on evaluation includes the analysis of the think-aloud data to identify areas of weakness. We aim to evaluate the specific aspects of CENTRIFUSER that adapt different librarian strategies and assess their effectiveness. In addition, we will carry out a web-based, large-scale evaluation that will focus on the ranking questions to reach statistical significance.

Our future work with respect to librarian techniques focuses on the information access mode of searching. We are concentrating on doing a finer grained analysis of indicative summaries as exemplified by annotated bibliography entries, which are more general and varied than subject guides and

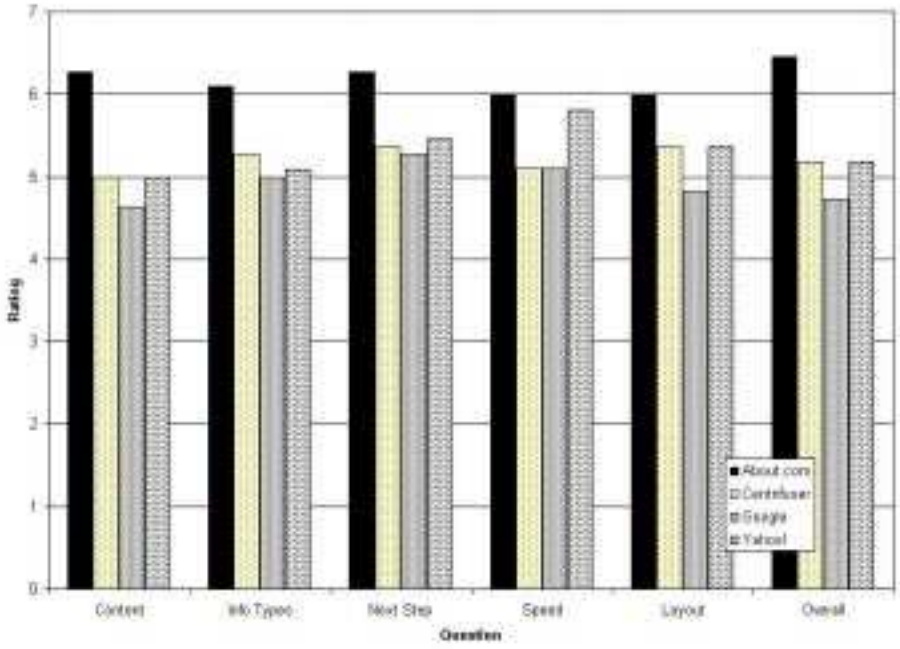

Figure 5: Preliminary evaluation ranking survey results. Results are not statistically significant.

the catalog entries as embodied by MARC catalog 520 fields. Our current research focuses on corpus-based natural language generation, which will enable the system to generate indicative summaries that mimic existing summaries in style and content.

10. CONCLUSIONS

We presented CENTRIFUSER, a IR post-processor that uses these observations to present an alternate visualization of the documents coming from a standard IR framework. To our knowledge this is the first such system which employs automatic text summarization specifically for post-processing and displaying results of an IR search. CENTRIFUSER employs text summarization to extract representative text across the documents that describes their similarities as well as to generate new text to capture their differences.

The different types of output in CENTRIFUSER's user interface were motivated by a user needs analysis based on our study of how two different types of librarians assist their users. We examined how reference librarians as well as cataloguers assist their users in the search and retrieval process in the physical library. From an analysis of their interaction strategies in both an interactive setting (with informational reference interviews) and an offline one (with subject guides and annotated bibliographies), we compiled an inventory of eight different strategies. Seven of these strategies have a direct impact on user interface design in IR, of which six are currently addressed by CENTRIFUSER's output.

Replicating these strategies in the online IR framework brought up differences in modality and attention span that influenced their nature. We found that since reference interviews are highly interactive and utilize many sources of modality, porting these strategies to an online IR setting requires more restrictions. On the other hand, static, offline strategies represented by subject guides and annotated bibliographies gain more flexibility when they are adapted to an online setting.

An initial evaluation of the system indicates several results that are promising. CENTRIFUSER's post-processing yields an improvement over the baseline user interface of presenting output in the standard IR framework user interface (as a ranked list of relevant documents). Results hint that fully automatic post-processed output approaches the same level of satisfaction for systems that require human intervention.

Text is a powerful output modality since it is easy to understand and is also a prerequisite for advanced user interfaces involving speech. CENTRIFUSER's ATS framework is based on topic structure, and allows results to be reported at different levels of granularity, resulting in longer or short summaries. This flexibility is being utilized in our current user interface project [17] that allows the summaries to be tailored lengthwise for different screen sizes (desktop screens, PDAs and cell phone displays) and allows contextual zooming on specific subtopics.

11. ACKNOWLEDGEMENTS

This research is supported by the National Science Foundation under Digital Library Initiative Phase II Grant Number IIS-98-17434. The authors would like to thank the staff of the Health Science Library at Columbia University, especially Elizabeth LaRue, for their time and input in suggesting supporting reference materials. We are also indebted to our evaluation collaborator, Dr. Andre Kushniruk of York University in helping design and supervise the project evaluation.

12. REFERENCES

[1] E. Amitay. *What Lays in the Layout: Using anchor-paragraph arrangements to extract descriptions of Web documents*. PhD thesis, Macquarie University, Sydney, Australia, Unpublished.

[2] ANSI. American national standard for describing books in advertisements, catalogs, promotional materials and book jackets. New York, USA, 1979.

[3] R. Balay, editor. *Guide to Reference Books*. American Library Association, Chicago, USA, 11th edition edition, 1996.

[4] M. J. Bates. Information search tactics. *Journal of the American Society for Information Science*, 30(7):205–214, 79.

[5] C. Borgman. Why are online catalogs hard to use? lessons learned from information-retrieval studies. *Journal of the American Society for Information Science*, 37(6):387–400, 1986.

[6] M. Chamlers and P. Chitson. BEAD: Exploration in information visualization. In *Proc. of 15th SIGIR '92*, 1992.

[7] J. Ding, L. Gravano, and N. Shivakumar. Computing geographic scopes of web resources. In *Proc. of the 26th Intl. Conf. on Very Large Data Bases*, 2000.

[8] R. Fidel. Writing abstracts for free-text searching. *Journal of Documentation*, 42(1):11–21, March 1986.

[9] V. Hatzivassiliglou, J. L. Klavans, M. L. Holcombe, R. Barzilay, M.-Y. Kan, and K. R. McKeown. Simfinder: A flexible clustering tool for summarization. In *Proc. of the Workshop on Automatic Summarization, NAACL 2001*, 2001.

[10] M. A. Hearst. Tilebars: Visualization of term distribution information in full text information access. In *Proc. of CHI 1995*, 1995.

[11] E. Z. Jennerich and E. J. Jennerich. *The Reference Interview as a Creative Art*. Libraries Unlimited, Littleton, Colorado, 1987.

[12] M.-Y. Kan, K. McKeown, and J. Klavans. Domain-specific informative and indicative summarization for information retrieval. In *Proc. of the Document Understanding Conference (DUC)*, pages 19–26, New Orleans, USA, 2001.

[13] W. A. Katz, editor. *Introduction to Reference Work*. McGraw-Hill, New York, USA, fifth edition edition, 1987.

[14] J. Koenemann and N. Belkin. A case for interaction: a study of interactive information retrieval behavior and effectiveness. In *Proceedings of CHI '96*, pages 205–212, Vancouver, Canada, May 1996. ACM Press.

[15] Library of Congress. Marc 21 format for classification data : including guidelines for content designation. Washington, D.C., USA, 2000. ISN 0660179903.

[16] E. Liddy. The discourse-level structure of empirical abstracts: An exploratory study. *Information Processing and Management*, 27(1):55–81, 1991.

[17] S. Lok and S. K. Feiner. The AIL automated interface layout system. In *Proc. of Intelligent User Interface*, 2002.

[18] I. Mani and E. Bloedorn. Summarizing similarities and differences among related documents. *Information Retrieval*, 1(1-2):35–67, 1999.

[19] K. McKeown, S.-F. Chang, J. Cimino, S. Feiner, C. Friedman, L. Gravano, V. Hatzivassiloglou, S. Johnson, D. Jordan, J. Klavans, A. Kushniruk, V. Patel, and S. Teufel. PERSIVAL, a system for personalized search and summarization over multimedia healtcare information. In *Proc. of the 1st JCDL*, Roanoke, USA, 2001.

[20] E. Morse, M. Lewis, and K. A. Olsen. Testing visual information retrieval methodologies case study: Comparative analysis of textual, icon, graphical and "spring" displays. *Journal of the American Society for Information Science and Technology*, 53(1):28–40, 2002.

[21] B. Nicolas J. Interaction with texts: Information retrieval as information-seeking behavior. *Information Retrieval*, pages 55–66, 1993.

[22] ODP. Open Directory Project guidelines. http://dmoz.org/guidelines.html, Last accessed November 2000.

[23] C. D. Paice. Constructing literature abstracts by computer: techniques and prospects. *Information Processing and Management*, 26(1):171–186, 1990.

[24] D. H. Sonnenwald. Developing a theory to guide the process of designing information retrieval systems. In *Proc. of the 15th SIGIR*, 1992.

[25] C. Witcombe. Art history resources on the web. http://witcombe.sbc.edu/ARTHLinks.html, Last accessed January 2002.