Evaluation of the Impact of User-Cognitive Styles on the Assessment of Text Summarization

Hien Nguyen, Eugene Santos, Jr., *Senior Member, IEEE*, and Jacob Russell

***Abstract*—Text summarization techniques have been found to be effective with regard to helping users find relevant information faster. The effectiveness and efficiency of a user's performance in an information-seeking task can greatly be improved if he/she needs to only look at a summary that includes the relevant information presented in his/her preferred manner. On the other hand, if the main idea is misrepresented and/or omitted altogether from a summary, it may take users more time to solve a target problem or, even worse, lead users to make incorrect decisions. There is an important need to design a personalized text summarization system that takes into account both what a user is currently interested in and how a user perceives information. The latter factor is referred to as a user's cognitive styles. Although there are some existing approaches that have employed a user's interests to help in the design of a personalized text summarization system, there has been inadequate focus on exploring cognitive styles. This paper aims at studying the impact of a user's cognitive styles when assessing multidocument summaries. In particular, we choose two dimensions of a user's cognitive style—the analytic/wholist and verbal/imagery dimensions—and study their impacts on how a user assesses a summary that was generated from a set of documents. In particular, the type of a document set refers to whether the set's content is loosely or closely related. We use a document set type to explore if there are any differences in the users' assessments of summaries that were generated from sets of different types. The results of this paper show that different users have different assessments with regard to information coverage and the way that information is presented in both loosely and closely related document sets. In addition, we found that the coherency ratings that were given to summaries from the two types of document sets were significantly different between the analytic and wholist groups. This result leads us to investigate the impact of a user's cognitive styles and the following two factors that directly relate to the coherence of a summary: 1) graph entropy and 2) the percentage of standalone concepts. We found that these two factors and a user's cognitive styles affect a user's ratings on the coherency of a summary.**

***Index Terms*—Cognitive styles, text summarization, user study, wholist/analytic.**

Manuscript received June 30, 3009; revised July 16, 2010; accepted September 9, 2010. Date of publication March 22, 2011; date of current version October 19, 2011. This work was supported in part by the University of Wisconsin (UW)–Whitewater Faculty Development Grant, the National Geospatial-Intelligence Agency under Grant HM1582-05-1-2042, the Intelligence Advanced Research Projects Activity (IARPA), and Securboration, Inc. This paper was recommended by Associate Editor J. A. Adams.

H. Nguyen is with the Department of Mathematical and Computer Sciences, University of Wisconsin–Whitewater, Whitewater, WI 53190 USA (e-mail: nguyenh@uww.edu).

E. Santos, Jr. is with the Thayer School of Engineering, Dartmouth College, Hanover, NH 03755 USA (e-mail: Eugene.Santos.Jr@Dartmouth.edu).

J. Russell is with the Department of Computer Sciences, University of Wisconsin–Milwaukee, Milwaukee, WI 53211 (e-mail: russel53@uwm.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCA.2011.2116001

I. Introduction

As BOTH offline and online information resources continue their explosive growth, the biggest challenge continues to be quickly finding relevant information from massive amounts of data. Text summarization has been an effective technique that is often used in combination with information retrieval and information-filtering applications to help users save time finding critically relevant information and making timely decisions [21], [26], [27]. Oftentimes, information retrieval applications use text summarization techniques to present users with the important points of a whole document or web page so that users can quickly scan and still precisely assess whether that document is related to their information needs [26]. Therefore, if a summary correctly presents the key information in a whole document, a user's information-seeking task will be more efficient. On the other hand, if important information is omitted or inadequately emphasized, a user may take more time to find the relevant information or, even worse, make an incorrect decision based on incomplete or incorrect information.

According to Mani and Maybury [26], text summarization is "the process of distilling the most important information from a source (or sources) to produce an abridged version for a particular user (or users) and task (or tasks)." A summary is thus a shorter version that contains the key information of a document or a set of documents. The latter approach is known as multidocument summarization. There are two primary inputs to any text summarization system: 1) documents and 2) users. The contents of the documents are used to produce the key points of the summary, whereas a user's interests or preferences are used to determine which points should be included or how they should be presented. This summarization process has been viewed as a function of the input documents and users [10]. User-modeling research is based on the idea that each individual has different ways of thinking and perceiving information and different ideas and interests on what should be considered as key points [8]. User-modeling techniques have widely been used to improve a user's performance in an information-seeking task (e.g., [3]). In the last few years, user models have been used in text summarization systems to improve summarization (e.g., [8]–[10], and [25]). The existing approaches capture and use a user's interests to enhance a text summarization system. Although a user's interests can help identify the commonality shared by a text summarization system and a user, other factors can influence their evaluation of a summary, including a user's cognitive styles or a user's stereotypes. A user's cognitive styles are used to describe the way each user thinks, perceives, and remembers information and are found to affect a user's information-seeking tasks [15]. Unfortunately, a background literature search turned up little in previous research on text summarization that has taken a user's cognitive styles into account. In addition, we lack results from comprehensive user studies that investigate the effects of a user's cognitive styles on the user's assessment of text summarization.

The main challenges for this type of study include choosing instruments and designing experiments while avoiding confounding. The instruments that were chosen need to bridge a user's cognitive styles and the process of assessing a summary. In this paper, we take advantage of well-established tests for cognitive styles from psychology and the links between research on a user's cognitive styles and areas related to the assessment of text summarization, e.g., information seeking and reading.

We explore how a user's cognitive styles can impact his/her assessment of multidocument summaries. The results reported here include the extended analysis of the earlier conference paper [32] that explores the effect of some key factors of a summary and a user's cognitive styles on the assessment of text summarization. This paper was initially based on the extension of a pilot experiment [52], in which five graduate students participated in evaluating 40 abstracts from four document sets in the Document Understanding Conference (DUC) 2002 collection. In the pilot study, we found that people were indifferent to the different types of document sets and tended to focus only on the general information covered in a summary. One goal of this paper is to verify whether this claim still holds with a much larger sample size. This paper also provides supporting data to verify whether and how a user's cognitive styles affects the process of assessing text summarization and, therefore, can benefit researchers in both the user modeling and text summarization communities.

We chose the following two dimensions of a user's cognitive styles that likely affect the process of assessing text summarization: 1) the analytic/wholist dimension and 2) the verbal/imagery dimension (see [35] and [36]). The analytic/wholist dimension measures the preferred way that a user perceives information in parts or in its entirety, whereas the verbal/imagery dimension measures the preferred way that a user processes information in textual description (verbal) or illustrations (imagery). These dimensions have widely been used to study the effect of a user's cognitive style on learning and information seeking, which closely relate to the process of assessing a summary (see [35]–[39]). We used four document sets from the DUC 2005 collection [7]. Thirty undergraduate students at the University of Wisconsin–Whitewater participated in this research work. All of these students went through the cognitive style analysis (CSA) test [35], [36], read all documents in these four document sets, and read and ranked summaries that were generated from ten randomly chosen summarization systems in the DUC 2005. We also made sure that these summaries were acceptable in terms of linguistic quality while maintaining the diversity of summaries. The results show that different users have different assessments with regard to information coverage and the way that information is presented in both loosely and closely related document sets. If the commonality of the concepts contained in any two documents of a document set is large, this set is labeled as a closely related document set. Otherwise, the set is referred to as loosely related.

Furthermore, we found that analytic, intermediate, and wholist groups are significantly different in their assessment of the coherence of summaries from loosely and closely related document sets. Wholists tend to process information as a whole, whereas analytics tend to process information in parts [35]. We also found that the wholist/analytic dimension, together with graph entropy and the percentage of stand-alone concepts in a document, affects a user's assessment of coherency of a summary.

This paper is organized as follows. In Section II, we provide some background about a user's cognitive styles and various personalized text summarization systems. Then, in Section III, we describe the experimental procedure and discuss the results in Section IV. Next, in Section V, we present a follow-up study that explores some potential factors that may help in automatically determining a user's cognitive styles. Finally, we present conclusions and future research possibilities.

II. Related Work

This paper focuses on exploring the effects of a user's cognitive styles on the assessment of text summarization. Therefore, in this section, we first provide an overview of some fundamental cognitive styles and then review some relevant studies that show the impact of a user's cognitive styles on learning and information seeking. Finally, we review some relevant work on personalized text summarization systems.

A person's cognitive style is viewed as his/her preferred way of thinking, perceiving, and remembering (for comprehensive reviews, please refer to [36] and [39]). Here, we choose to discuss the following four different dimensions of a user's cognitive styles that have been found to affect learning in both offline and online settings, as well as a user's preferences in using virtual environments, during web browsing, and while searching:

1) holist/serialist [33];
2) field dependence/independence [50];
3) wholist/analytic [36];
4) verbal/imagery [36].

The holist/serialist dimension is defined by Pask [33], with a focus on learning. Holists tend to use a *global* approach to learning, whereas serialists tend to narrowly or *locally* concentrate on the details of the topics learned [22]. This dimension can be measured by a number of different tests, e.g., the study preference questionnaire [12].

The field dependence/independence dimension [50] measures the degrees "to which a learner's perception or comprehension of information is affected" [22, p. 87] by the surrounding environment or fields. Field-dependent people may find it hard to find the information that they are looking for, given the noise and ill-defined problems that they are working on, whereas field-independent people can find ways of recognizing the relevant information or making problems that they are working on more concrete. Field-dependent people focus on the global dimensions of a problem, whereas field-independent people focus on the details of a problem.

The wholist/analytic dimension [36] is closely related to field dependence/independence [50]. This dimension reflects the preferred way that a user organizes or processes information either in its entirety (wholist) or in parts (analytic). Analytic users may have difficulty seeing the big picture when solving a problem, whereas wholist users may have difficulty decomposing a complicated problem into smaller subcomponents. Riding and Rayner [39] have shown that the wholist/analytic dimension can be mapped to the field dependence/independence dimension. The mapping is determined as follows: we use the CSA test (see [35] and [36]) to measure a user's cognitive style. If the number obtained at the end of the test is below 1.03, it is equivalent to field-dependent individuals, whereas if it is greater than 1.36, it is equivalent to field-independent individuals.

In addition to the wholist/analytic dimension, Riding and Rayner [35] measures a user's cognitive style by the verbal/imagery dimension. Some researchers have found that this dimension affected a user's learning. For example, verbalizers tend to prefer the structured–verbal type of materials, whereas imagers tend to prefer the structured–pictorial types of materials [37], [38]. Another study on impacts of this dimension of a user's cognitive style on web-based instructional systems [20] shows that verbalizers outperform imagers in essay-type problems in both short- and long-page conditions. In general, verbalizers are better at working with verbal information, whereas imagers are better at keeping track of where they are in a web-based system. More interestingly, also in the web-based learning domain, Cook [6] has found that imagers are likely to stick with the tasks that they find boring, whereas verbalizers need stimulating presentations of a task. Last, the verbal/imagery dimension has been found to affect the users' preferences for clip dynamics, frame rate, and colors in video-based applications [17], [18].

These four dimensions (holist/serialist, field dependent/independent, wholist/analytic, and verbal/imagery) of a user's cognitive styles essentially address the global–local issue of a user's preferences while performing information seeking, analyzing information, and solving problems. Most of the existing studies have investigated the impacts of a user's cognitive styles on browsing/searching and online learning. Some interesting user studies have shown that a user's wholist/analytic dimension affects the choice of search strategies and retrieval performance, including the use of Boolean operators, changes in search strategies such as browsing and searching, and the assessment of document relevancy (e.g., [13], [15], and [49]). In online learning, several studies have found that the field dependence/independence dimension affects learning preferences and navigation behaviors in hypermedia systems (e.g., [5] and [24]). A comprehensive review described by Chen and Macredie (2002) [5] showed that 30 studies have found links between field-dependent individuals and the factors that affect a learner in a hypermedia system, e.g., their navigations in hyperspace, disorientations, matching and mismatching, learning modes, and learning effectiveness. Last, another major emphasis was to explore the impact of a user's cognitive styles on how graphical user interfaces are used (see [5], [16], and [19]). Factors such as user controls and the choices of different options have been found to be affected by the field dependence/independence dimension.

As aforementioned, humans and documents are two critical inputs to any text summarization technique. Therefore, it is also intuitive to use some user modeling techniques to enhance the summarization systems. Personalizing a summary often contains the following two steps: 1) determining a user's topical interests and 2) constructing the summary surrounding those interests using heuristics (see [1], [4], and [9]). A user can provide a text summarization system with a list of topics that describes his/her current interests, or the system can infer user interests by extracting the user query's terms or the most frequently used keywords from the documents that the user is looking at. For example, the text summarization approach described by Maña *et al.* [25] generated a personalized summary based on a set of words that were derived from the user's query together with related words found in the Wordnet taxonomy [28]. After a user's interests have been determined, a personalized summary is generated using the following three heuristics: 1) location-based heuristics (e.g., important terms in a title or terms at the beginning of a paragraph); 2) linguistics-based heuristics (e.g., certain phrases); and 3) statistics-based heuristics (e.g., frequency). For example, query terms and term positions have been used by Tombros and Sanderson [48] to generate summaries. Another example of linguistic- and statistics-based approaches is the multilayer model [9], in which each topical interest is associated with a corresponding weight. Then, the vector space model is applied to compute the cosine similarities between each sentence of a document and the interest vector. A summary is generated by selecting the top 20% of ranked sentences. Another multitier model is described by Díaz and Gervás [8].

Text summarization systems that use user-modeling techniques to generate a personalized summary surrounding a user's interests have been preferred by human users throughout various evaluations [6], [8]–[10]. A user's satisfaction with a search enhanced by a personalized text summarization technique is improved compared to a search with generic summaries [9]. However, some key problems are still not addressed. First, the relationships among the keywords added to a user's interests list are still ignored. Second, such techniques take into account *what* a user is currently interested in searching or learning but not *how* he/she goes about doing a certain task. In particular, a user's preferred styles of reading, memorizing, and recalling events are largely omitted. Assessing a summary directly involves reading, processing, recalling, and presenting information. This condition leads us to performing the study in this paper, in which we explore the possible link between a user's cognitive styles and the process of assessing text summarization. Additionally, a user's cognitive styles have been found to affect all of the subprocesses of assessing a summary (e.g. reading [40], learning [51]). We hope to highlight the relationships between a user's cognitive styles and evaluating text summarization systems. The results in this paper can also help researchers in further exploring possible relations between a user's characteristics and behaviors and a set of documents to successfully design and evaluate a personalized text summarization system.

III. Experimental Procedure

A. Objectives, Cognitive Style, Instruments, and Hypotheses

Objectives: We aimed at identifying the impacts of a user's cognitive styles in assessing a summary in multidocument summarization. In particular, we created an experimental procedure and collected data to answer the following three questions.

1) Are users' assessments of multidocument summarization significantly different with regard to different types of document sets?
2) How do a user's cognitive styles impact his/her assessment of text summarization?
3) What are some factors in the content of a summary that can help us automatically detect a user's cognitive styles?

The motivation and basis behind these objectives are threefold. First, as aforementioned, in [52], we found that people's ratings were indifferent to the type of document sets. We want to verify if this result still holds with a larger sample size. By the "type" of a document set we refer to its degree of related content, i.e., whether the content of a document set is loosely or closely related. Second, as previous research results show, a user's cognitive styles affect how a user conducts reading, information seeking, and learning. We would like to find similar evidence to show that a user's cognitive styles also affect the interpretation of the text summarization process. Last, if a user's cognitive styles affect any parts of text summarization, it would be useful and necessary for a text summarization system to *automatically* detect a user's cognitive styles by analyzing the user's interactions with summaries. By doing so, we can avoid having users take a separate cognitive test every time they use a text summarization system.

Cognitive Style: In this paper, we have focused on the following two dimensions of a user's cognitive style: 1) wholist/analytic and 2) verbal/imagery. These two dimensions have extensively been studied by Riding and his collaborators (see [35]–[39]) and have widely been used to study the effect of a user's cognitive styles on learning and information seeking. The first dimension, wholist/analytic, measures whether a user prefers to process information by looking at its entirety or by decomposing information into smaller pieces. The second dimension, verbal/imagery, reflects whether a user mentally perceives, recognizes, and organizes either in words or in a graphical representation. Verbal users are more comfortable with verbally expressive tasks, whereas imagery users prefer information and tasks that are visually presented.

Instruments: The following two main instruments were used for this paper: 1) the CSA test [35], [36] and 2) the DUC 2005 document collection. The CSA test is a computerized tool that is used to measure the wholist/analytic and verbal/imagery dimensions of a user's cognitive style. This test measures these two dimensions by comparing the response time of a user as he/she responds to an image/verbal or analytic/wholist question. This test consists of three parts. Each part contains multiple-choice questions on whole–parts and set–subsets. Each participant spent about 7–10 minutes to finish this test. At the end of the test, with regard to the wholist/analytic dimension, each user was assigned to one of the following three groups, depending on his/her score: 1) wholist; 2) analytic; or 3) intermediate. With regard to the verbal/imagery dimension, each user was assigned to one of the following three groups: 1) verbal; 2) bimodal; or 3) imager.

We chose four collections from the DUC 2005 that satisfy the following two categories: 1) the main topics of these collections were interesting so that a young audience would be more likely to participate in this study and 2) these collections were different in terms of the *type* of document sets. In particular, they have different degrees of related content. The process of measuring this degree is described as follows. We focus on the type of document sets, because we would like to further extend the results of the pilot study described in [52].

Step 1. Given a document set $D = \{d_1, d_2, d_3, \ldots d_n\}$, we convert each document to a document graph (DG), which is a directed acyclic graph that contains concept and relation nodes (see [30], [31], [41]–[43], [45], and [46]). Concept nodes represent noun phrases (NPs), whereas relation nodes describe relationships among the concept nodes. We currently support the following two kinds of relation nodes: 1) "*is a*" and 2) "*related to*." The construction of a DG is an automated process, which contains the following steps.

- Tokenize a document in plain text format into sentences.
- Parse each sentence using Link Parser [47].
- Extract NPs from the parsing results.
- Generate relations between concepts/entities based on heuristic rules and put them into a graph format.

Fig. 1(b) shows a small part of a DG that was generated from a document about industrial spies in Volkswagen and General Motors in the DUC 2005 collection. The content of this document is shown in Fig. 1(a). The terms highlighted in this figure are shown in the corresponding DG in Fig. 1(b). DG representation has been used and found to be effective in improving a user's performance in information seeking (see [30], [31], [41]–[43], [45], and [46]). Please see Appendix [41] for the detailed heuristics.

Step 2. The degree of related content of this document set is computed with the following formula:

$$D = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n}\sum_{j=1,j\neq i}^{n} sim(d_i, d_j)$$

in which

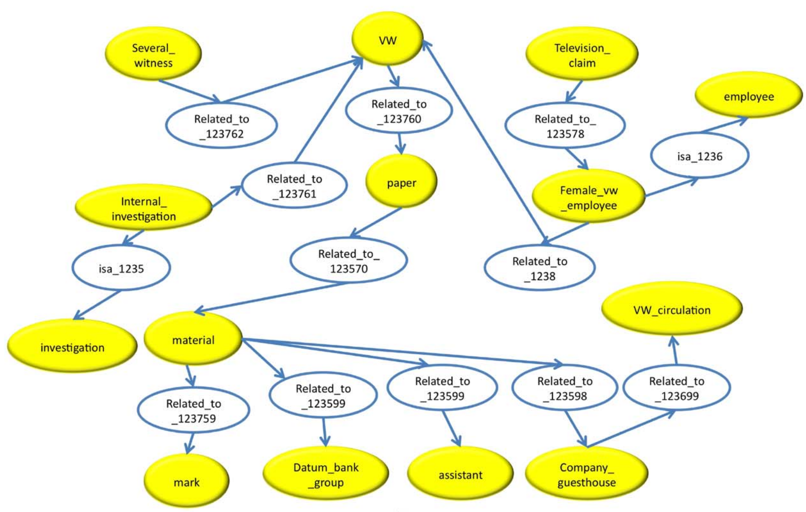
$$sim(d_i, d_j) = \frac{|C_i \cap C_j|}{|C_i \cup C_j|} + \frac{|R_i \cap R_j|}{|R_i \cup R_j|}.$$

In this equation, C_i, C_j are the sets of concepts in d_i, d_j, and R_i, R_j are the set of relations in d_i, d_j. Two relation nodes are matched only when at least one of its parent nodes and one of its child nodes are both matched.

One example of a group of closely related documents is set $D311$ in the DUC 2005 collection, which consists of 41 documents about industrial spying between Volkswagen and General Motors. These articles are about the events that occurred from 1993 to 1994. These events involve specific people, locations, timelines, and actions. Such factors are repeated throughout the

FT 13 AUG 93 / VW probe may recall witnesses
PUBLIC prosecutors investigating spying and theft allegations against. Volkswagen employees may recall witnesses for further questioning, following reports that data belonging to Adam Opel, the German subsidiary of General Motors, may have been punched into VW computers. The possibility arose yesterday after a televised claim that a female VW employee had said she and nine colleagues had been instructed to store material by an assistant to Mr Jose Ignacio Lopez de Arriortua, VW's new production director, and the man at the centre of the investigation. The woman, allegedly already interrogated, had apparently said nothing about the origins of the material when questioned. However, she reportedly told a third party that the documents carried an Opel logo. Mr Georg Nauth, a senior prosecutor and spokesman for the investigators, said the reports would be checked. It is understood several witnesses from VW, under questioning, have said they could not remember - or could not rule out - whether material they handled bore marks identifying their origins as either Opel or GM. Mr Walter Hiller, a member of the VW supervisory board, said yesterday the board had been 'credibly assured' at a meeting last Friday that an internal investigation at VW found there was no such material in the group's data banks.
The meeting unanimously backed Mr Lopez and said there was no evidence to warrant accusations of industrial espionage. Meanwhile, Opel threatened further legal action if VW did not tell it in precise detail the nature and contents of material destroyed in the week of March 22 on instructions of Mr Lopez. VW last weekend admitted that papers, including possibly secret or sensitive material, were destroyed at its company guesthouse to prevent its circulation within VW.

(a)



(b)

Fig. 1. (a) Content of a document. (b) Small part of a DG generated from that document.

collection. One example of loosely related documents is set $D354$, which consists of 46 documents about the activities of various journalists overseas. This collection covers different regions, e.g., America, Europe, Asia, the Middle East, and Africa. It also covers different activities of interests, e.g., politics, the Soviet–Afghanistan war, Muslim war, economic issues, and United Nations activities.

Hypotheses: To meet the first two objectives, we have constructed the following hypotheses.

1) *Users' informativeness and coherency ratings for different types of document sets are significantly different.* Intuitively, it should be much more difficult to create a good summary for the loosely related document sets. This means that users may prefer summaries that were generated by closely related document sets. However, the pilot study conducted in 2005 with five graduate assistants in [52] found that all five participants' ratings are indifferent to the summaries that were generated from different types of document sets. We would like to see if this result still holds for a much larger data sample of participants.
2) *Wholist users will prefer summaries that were generated from closely related document sets, whereas analytic users will prefer summaries that were generated from loosely related document sets.* Wholists tend to process information as a whole, whereas analytics tend to process information as parts [36]. In other words, it would be easier for wholist users to process a summary if it focuses on one common viewpoint, whereas it is easier for analytic

users to process a summary if it includes some additional details. The similarities of the documents in the closely related document set are higher than the corresponding ones from the loosely related document sets. Therefore, the probability of picking up the concepts of the same common topic to a summary from the closely related document sets is higher than from the loosely related document sets.

3) *Verbalizers will prefer summaries that were generated from a loosely related document set.* Verbalizers prefer summaries that contain more detailed information over summaries that contain less detailed information. The basis here is that verbal users learn best from word and verbal description [6], and therefore, the more specific detail descriptions a summary contains, the higher the ratings that verbalizers may give a summary. The contents of documents in a loosely related document set are less overlapped compared to the contents in a closely related document set. A summary that was generated from a loosely related document set may contain different concepts that were extracted from the documents in this set. Therefore, such a summary may contain more detailed information compared to the summary that was generated using the same technique from a closely related document set.

The third objective will be addressed by using the results of the experiment in this section and by exploring some attributes that relate to the coherence and informativeness of a summary. We will discuss this objective in detail in Section V.

B. Testbed

The following four document sets from the DUC 2005 collection have been chosen for this paper:

1) robot technology;
2) journalism overseas;
3) industrial spies in Volkswagen and General Motors;
4) tobacco companies.

We chose these document sets, because their main topics, e.g., technology, spies, and tobacco, were exciting and likely to be attractive to the young participants of this paper. In addition, the two collections on journalism overseas and robot technology are loosely related document sets ($D = 0.032$ and $D = 0.029$, respectively). The remaining two sets on industrial spies in Volkswagen and General Motors and tobacco companies are closely related document sets ($D = 0.085$ and $D = 0.084$, respectively). For each of these aforementioned document sets, we chose ten summaries that were generated from ten summarization systems numbered in the DUC 2005 collection as 1, 11, 12, 13, 14, 15, 16, 17, 18, and 19. These summaries were acceptable in terms of linguistic quality, which includes the following five categories:

1) grammaticality;
2) nonredundancy;
3) referential clarity;
4) focus;
5) structure and coherence.

By acceptable we mean that none of the summaries has been rated as "very poor" by assessors from the National Institute of Standard and Technology (NIST) in any of these five categories. One example of an unacceptable summary is a summary that is grammatically correct but does not have any main focus or any structure. Another example is a summary that contains no redundancy but is grammatically incorrect and is incoherent.

Thirty undergraduate students were recruited from the University of Wisconsin–Whitewater. These students were majoring in mathematics, social work, management computer systems, and information technology infrastructure. Each student was paid $30 upon completion of the study. There were 15 female and 15 male participants, with their ages ranging from 17 to 44 years old.

C. Procedure

Step 1. Each participant took the CSA test to determine his/her wholist/analytic and verbal/imagery cognitive style.

Step 2. Each participant was given the aforementioned four sets of documents and 40 summaries. The participants received detailed instructions on the whole process. In particular, they were first required to carefully read each document set and identify the most important information from that set. Then, they were asked to evaluate the summaries for that corresponding set by assigning two scores to each of the summaries using a five-point-scale scoring system (1 is good, and 5 is bad). One score represents the user's informativeness rating, and the other score represents the user's coherency rating. The informativeness rating is the participant's assessment of how well the information has been covered, whereas the coherency rating is the participant's assessment on how well the summary has been written. We orally walked the participants through an example and thoroughly explained the two aspects. One example of an abstract with written rating instruction is shown in the Appendix.

The order of the document sets that were given to each participant was changed to avoid the effects of fatigue near the end of the process on the outcomes of the study. The four document sets were numbered from 1 to 4, and all possible combinations of these four numbers were generated. There were 24 different orders. We then generated a random number that corresponded to one of these orders and assigned them to a participant. No more than two participants had the same order of collections. If each participant was given the document sets in the same order, the collected data might be skewed, because the collection rated last might get false ratings from all participants. By changing the order of document sets, we experimentally balanced the effect of fatigue on outcomes.

D. Choice of Statistics Tools

We chose the one-way analysis of variance (ANOVA) tool to find whether the differences in terms of informativeness and coherency ratings of different document sets were significant. ANOVA is a statistical technique that is widely used for comparing the means of two or more samples [11]. The

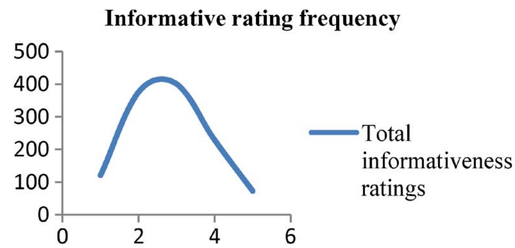


Fig. 2. Total informativeness ratings by informativeness rating values.

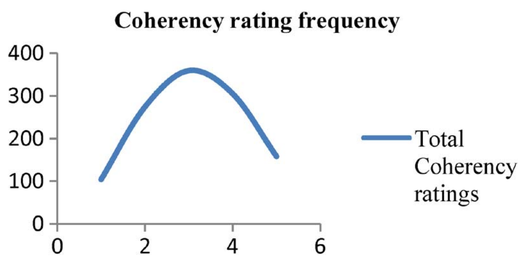


Fig. 3. Total coherency ratings by coherency rating values.

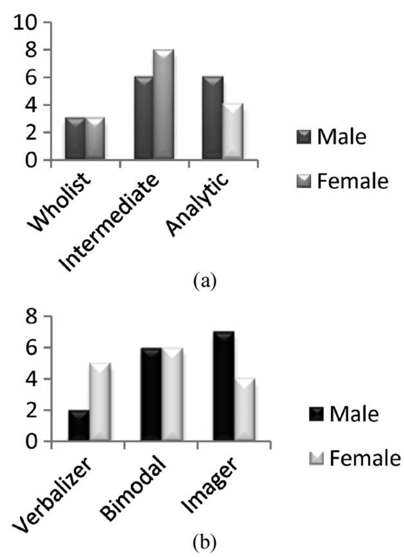


Fig. 4. (a) Wholist/Analytic dimension of participants by gender. (b) Verbal/Imagery dimension of participants by gender.

informativeness and coherency ratings collected satisfy the prerequisites of this analysis tool, because both ratings are approximately normally distributed. The frequencies of informativeness and coherency ratings are shown in Figs. 2 and 3. The majority of ratings are clearly clustered around the means.

We chose a general linear model to analyze whether differences in users' cognitive styles affect users' ratings for different types of document sets, because the number of subjects varies between style groups.

IV. Analysis of the Results

A. Cognitive Styles of Participants

The CSA test measures the following two dimensions of a user's cognitive styles: 1) wholist/analytic and 2) verbal/imagery. The results for each dimension are accordingly broken down in Fig. 4(a) and (b). Note that the number of female verbalizers is more than double the number of male verbalizers, whereas the number of male imagers is more than the number of female imagers. As shown in Fig. 5, the dimension with the most participants is the intermediate–imager (two males and four females), and the dimension with the fewest participants is the wholist–verbalizer (no male and one female).

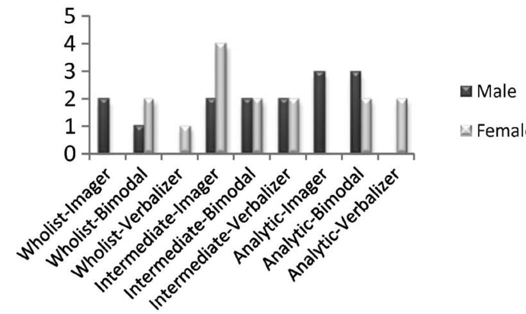


Fig. 5. User's cognitive styles by gender.

B. Hypothesis 1

Users' informativeness and coherency ratings for different types of document sets are significantly different.

The pilot study [52] suggested that there were no significant differences in informativeness ratings and coherency ratings over two different types of document sets. However, that study was done with only five participants. Similarly, as reported in the work of Nguyen *et al.* [32], we performed one-way ANOVA using the *average* ratings for 30 participants and found no significant differences. Therefore, in the analyses for this paper, we would like to see if the previous results hold if we use the ratings for each summary to acquire more precise results.

In the first analysis, the informativeness rating is considered a dependent variable. The independent variable is the type of document sets with two values (loosely and closely related document sets). The averages of informative ratings are different for loosely related ($\mu = 2.8817$) and closely related ($\mu = 2.7133$) document sets. We ran one-way ANOVA using SPSS [11] to see if this difference is significant. The result is shown in Table I. We found that there is a significant difference between the ratings of two groups (loosely and closely related document sets; Sig $= 0.006$, as highlighted in Table I).

Similarly, in the second analysis, coherency ratings are considered a dependent variable. Again, the averages of coherency ratings are different for loosely related ($\mu = 3.22$) and closely related ($\mu = 3.0117$) document sets. The result of one-way ANOVA run using SPSS is shown in Table II. There is a significant difference between the ratings of two groups (loosely and closely related document sets; Sig $= 0.002$, as highlighted in Table II).

By using the informativeness and coherency ratings for each summary, we found significant differences between the two types of document sets. Therefore, different users give significantly different ratings to the different types of document sets. This result is different from the pilot study [52] and preliminary analysis [32]. The most likely reasons are that, in the pilot study, we used a small sample size, and in the preliminary analysis, we

TABLE I
ONE-WAY ANOVA ON INFORMATIVENESS RATINGS OVER TWO TYPES OF DOCUMENT SETS

	N	Mean	Std. Dev	Std. Error
Loose	600	2.8817	1.08303	.04421
Close	600	2.7133	1.01795	.04156
Total	1200	2.7975	1.05392	.03042

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	8.501	1	8.501	7.696	.006
Within Groups	1323.292	1198	1.105		
Total	1331.793	1199			

Loose: loosely related document set.
Close: closely related document set.

TABLE II
ONE-WAY ANOVA ON COHERENCY RATINGS OVER TWO TYPES OF DOCUMENT SETS

	N	Mean	Std. Dev	Std. Error
Loose	600	3.2200	1.17213	.04785
Close	600	3.0117	1.13740	.04643
Total	1200	3.1158	1.15911	.03346

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	13.021	1	13.021	9.762	.002
Within Groups	1597.878	1198	1.334		
Total	1610.899	1199			

Loose: loosely related document set.
Close: closely related document set.

ran one-way ANOVA on the *averages* of all ratings over the two sets instead of using each rating for each summary.

TABLE III
DESIGN OF ANALYSIS BASED ON THE ANALYTIC/WHOLIST DIMENSION

		Type of document set	
		Loosely related sets	Closely related set
Cognitive style	Analytic	n= 200	n=200
	Intermediate	n= 280	n=280
	Wholist	n= 120	n=120

TABLE IV
RESULTS OF THE GENERAL LINEAR MODEL UNIVARIATE FOR INFORMATIVE RATINGS OVER COGNITIVE STYLES

Descriptive Statistics

Dependent Variable:InformativenessRating

CognitiveStyle	Type	Mean	Std. Dev	N
Wh	Loose	2.90	1.088	120
	Close	2.59	1.017	120
	Total	2.75	1.062	240
Ana	Loose	3.04	1.072	200
	Close	2.85	1.066	200
	Total	2.94	1.072	400
Int	Loose	2.76	1.078	280
	Close	2.67	.976	280
	Total	2.72	1.029	560

Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	22.837[a]	5	4.567	4.166	.001
Intercept	8355.280	1	8355.280	7621.498	<.001
CognitiveStyle	12.315	2	6.158	5.617	.004
Type	10.338	1	10.338	9.430	.002
CognitiveStyle * Type	2.020	2	1.010	.922	.398
Error	1308.956	1194	1.096		
Total	10723.000	1200			
Corrected Total	1331.792	1199			

Wh: wholist; Ana: analytic; Int: intermediate

C. Hypothesis 2

Wholist users will prefer summaries that were generated from closely related document sets, whereas analytic users will prefer summaries that were generated from loosely related document sets. (In other words, wholist users will give better ratings to the summaries that were generated from closely related document sets, whereas analytic users will give better ratings to summaries that were generated from loosely related document sets.)

In this analysis, we have the following two independent variables: 1) cognitive style and 2) type of document sets. The cognitive-style variable has the following three values: 1) analytic; 2) intermediate; and 3) wholist. On the other hand, the type of document sets has the following two variables: 1) loosely related and 2) closely related. There is one dependent variable, which is the actual rating for each summary. This design is used for both informativeness and coherency ratings, respectively. Each document set has ten summaries, and there are four document sets. Therefore, we have 40 informativeness rating values and 40 coherency ratings values for each user. The design is shown in Table III.

We have the following three groups of users: 1) the analytic group with ten users; 2) the intermediate group with 14 users; and the 3) wholist group with six users. We first perform the general linear model univariate analysis using informativeness ratings. The result is shown in Table IV. We found that the difference in informativeness ratings by the cognitive-style group is significant ($\text{Sig} = 0.004$) and the difference in informativeness ratings by the type of document set is also significant ($\text{Sig} = 0.002$). However, we found no significant relationship between different groups of users by cognitive styles and the different types of document sets ($\text{Sig} = 0.398$). This result is also in line with the preliminary analysis [32], in which we performed two-way ANOVA using the average informativeness ratings for each group.

TABLE V
RESULT OF THE POST HOC TEST FOR COGNITIVE STYLES AND INFORMATIVENESS RATINGS

(I) CognitiveStyle	(J) CognitiveStyle	Mean Difference (I-J)	Std. Error	Sig.
Wh	Ana	-.1942	.08549	.060
	Int	.0280	.08078	.936
Ana	Int	.2221	.06854	.004

TABLE VI
GENERAL LINEAR MODEL UNIVARIATE ANALYSIS USING COHERENCY RATINGS OVER THE WHOLIST/ANALYTIC DIMENSION

Descriptive Statistics

Dependent Variable:CoherencyRating

CognitiveStyle	Type	Mean	Std. Dev	N
Wh	Loose	3.5833	1.22016	120
	Close	2.8750	1.22002	120
	Total	3.2292	1.26821	240
Ana	Loose	3.2100	1.19711	200
	Close	3.1800	1.20201	200
	Total	3.1950	1.19815	400
Int	Loose	3.0714	1.10160	280
	Close	2.9500	1.03920	280
	Total	3.0107	1.07162	560

Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	44.036[a]	5	8.807	6.711	<.001
Intercept	10531.586	1	10531.58 6	8025.40 7	<.001
CognitiveStyle	11.778	2	5.889	4.487	.011
Type	21.863	1	21.863	16.661	<.001
CognitiveStyle * Type	19.238	2	9.619	7.330	.001
Error	1566.863	1194	1.312		
Total	13261.000	1200			
Corrected Total	1610.899	1199			

a. R Squared = .027 (Adjusted R Squared = .023)
Wh: wholist; Ana: analytic; Int: intermediate

Furthermore, in the post hoc test for the cognitive-style variable (shown in Table V), we found that the analytic group is significantly different from the intermediate group, and vice versa $(\mathrm{Sig} = 0.004)$, as highlighted in Table V.

Similarly, we perform the general linear model univariate analysis using coherency ratings (shown in Table VI). The difference of coherency ratings by the cognitive-style group is significant $(\mathrm{Sig} = 0.011)$, and the difference of coherency ratings by the type of document set is also significant $(\mathrm{Sig} < 0.001)$. More importantly, there is a significant relationship between different groups of users by cognitive styles and different types of document sets $(\mathrm{Sig} = 0.001)$. This result is in line with [32], in which the two-way ANOVA using the average coherency ratings for each group was performed.

TABLE VII
DESIGN OF ANALYSIS BASED ON THE VERBAL/IMAGERY DIMENSION

		Type of document set	
		Loosely related document sets	Closely related document set
Cognitive style	Verbal	n=140	n=140
	Bimodal	n=240	n=240
	Imagery	n=220	n=220

D. Hypothesis 3

Verbalizers will prefer summaries that were generated from a loosely related document set. (In other words, verbalizers will give better ratings to the summaries that were generated from loosely related document sets.)

Similar to the design of the analysis done in Section IV-C, we used the following two independent variables for this analysis: 1) cognitive style and 2) type of document sets. We consider the verbal/imagery dimension for the cognitive style in this experiment. Therefore, the cognitive-style variable has the following three values:1) verbal; 2) bimodal; and 3) imagery. On the other hand, the type of document sets has the following two variables: 1) loosely related document sets and 2) closely related document sets. The rating of each summary is considered a dependent variable. This design is used for both informativeness and coherency ratings. Similar to the design in Section IV-C, we have 40 summaries for four document sets and, therefore, 40 informativeness rating values and 40 coherency ratings values for each user. The design of this analysis is shown in Table VII.

The verbal group has 7 users, the bimodal group has 12 users, and the imagery group has 11 users. We first perform the general linear model univariate analysis using informativeness ratings. The result is shown in Table VIII.

The difference in informativeness ratings by the verbal/imagery dimension of a user's cognitive style is significant $(\mathrm{Sig} < 0.001)$, and the difference in informativeness ratings by the type of document set is also significant $(\mathrm{Sig} = 0.022)$. However, we found no significant relationship between different groups of users by the verbal/imagery dimension of a user's cognitive styles and different types of document sets $(\mathrm{Sig} = 0.095)$. This result is also in line with [32], in which we performed two-way ANOVA using the average informativeness ratings for each group.

We performed a post hoc test for the verbal/imagery dimension of a user's cognitive styles. In the post hoc test for the cognitive-style variable (shown in Table IX), we found that the verbal group is significantly different from the bimodal group, and vice versa $(\mathrm{Sig} = 0.049)$. In addition, the imagery group was significantly different from the bimodal group, and vice versa $(\mathrm{Sig} < 0.001)$.

Next, we performed the general linear model univariate analysis using coherency ratings (as shown in Table X). The difference of coherency ratings by the verbal/imagery dimension of a user's cognitive-style group was not significant $(\mathrm{Sig} = 0.258)$, but the difference of coherency ratings by the type of document set is significant $(\mathrm{Sig} = 0.001)$. Last, we found that there is no significant relationship between the verbal/imagery dimension of a user's cognitive styles and different types of

TABLE VIII
GENERAL LINEAR MODEL UNIVARIATE ANALYSIS USING INFORMATIVENESS RATINGS OVER THE VERBAL/IMAGERY DIMENSION

Descriptive Statistics

Dependent Variable:InformativeRating

CognitiveStyle	Type	Mean	Std. Dev	N
Ver	Loose	2.74	1.108	140
	Close	2.79	1.035	140
	Total	2.77	1.071	280
Img	Loose	2.80	1.055	220
	Close	2.50	.894	220
	Total	2.65	.988	440
Bi	Loose	3.04	1.078	240
	Close	2.86	1.083	240
	Total	2.95	1.083	480

Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	35.215[a]	5	7.043	6.486	<.001
Intercept	8832.287	1	8832.287	8133.5 33	<.001
CognitiveStyle	21.586	2	10.793	9.939	<.001
Type	5.686	1	5.686	5.236	.022
CognitiveStyle * Type	5.129	2	2.564	2.361	.095
Error	1296.577	1194	1.086		
Total	10723.000	1200			
Corrected Total	1331.792	1199			

a. R Squared = .026 (Adjusted R Squared = .022)
Ver: verbalizer; Img: imager; Bi: bimodal

TABLE IX
POST HOC ANALYSIS FOR THE VERBAL/IMAGERY DIMENSION OF A USER'S COGNITIVE STYLES USING INFORMATIVENESS RATINGS

(I) CognitiveStyle	(J) CognitiveStyle	Mean Difference (I-J)	Std. Error	Sig.
Ver	Img	.1201	.07966	.287
	Bi	-.1842	.07836	.049
Img	Bi	-.3044	.06878	<.001

Ver: verbalizer; Img: imager; Bi: bimodal

document sets ($\text{Sig} = 0.360$). This result is also in line with the previous preliminary analysis [32], in which we performed two-way ANOVA using the average coherency ratings for each group.

V. EXPLORING FACTORS THAT HELP AUTOMATICALLY DETERMINE A USER'S COGNITIVE STYLES

A. Determining Relevant Factors

After finding that different users give significantly different coherency ratings for summaries that were generated from loosely and closely related document sets, we would like to determine a way of automatically detecting a user's cognitive styles by analyzing related content factors. We used this approach to achieve the third and final research objective of this paper. To help this process, we investigated the possible links between a user's cognitive styles and the factors that relate to coherence of a document. The coherence of a summary means that the concepts are related to one another so that readers can easily follow them (http://www.boisestate.edu/wcenter/ww97.htm). Because each summary is represented in a DG, coherence can be measured with the *connectivity* of a DG. Two factors for measuring this connectivity attribute are given as follows: 1) graph entropy and 2) the percentage of concepts that do not connect to any other concepts (referred to as stand-alone concepts in this paper). Graph entropy focuses on the structure of the graph as a whole and is considered a global measure [29]. High entropy values mean that several vertices are equally important, whereas low entropy values mean that there are few key concepts for this document. We define the percentage of stand-alone concepts as the ratio between the total of concept nodes that do not connect to any other concept nodes over the total of concept nodes for a summary.

Both measures directly relate to the coherence attribute of a summary. The overall goal of this experiment is that, if any of these factors relate to a user's cognitive styles, we can develop an automated algorithm to determine the wholist/analytic dimension of a user's cognitive styles through a document's content and a user's actions on the document content.

TABLE X
GENERAL LINEAR MODEL UNIVARIATE ANALYSIS USING COHERENCY RATINGS OVER THE VERBAL/IMAGERY DIMENSION

Descriptive Statistics

Dependent Variable:CoherencyRating

CognitiveStyle	Type	Mean	Std. Deviation	N
Ver	Loose	3.25	1.230	140
	Close	2.96	1.137	140
	Total	3.10	1.191	280
Img	Loose	3.20	1.112	220
	Close	2.91	1.088	220
	Total	3.05	1.108	440
Bi	Loose	3.22	1.196	240
	Close	3.13	1.175	240
	Total	3.18	1.185	480

Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	19.368[a]	5	3.874	2.906	.013
Intercept	10997.790	1	10997. 790	8250.7 75	.000
CognitiveStyle	3.620	2	1.810	1.358	.258
Type	14.002	1	14.002	10.505	.001
CognitiveStyle * Type	2.727	2	1.364	1.023	.360
Error	1591.531	1194	1.333		
Total	13261.000	1200			
Corrected Total	1610.899	1199			

a. R Squared = .012 (Adjusted R Squared = .008)
Ver: verbalizer; Img: imager; Bi: bimodal

B. Hypotheses

We hypothesize that the wholist/analytic group, together with graph entropy and the percentages of stand-alone nodes, will affect the user's coherency ratings.

The basis for this hypothesis is that wholist users significantly gave high coherency ratings for summaries that were generated from a closely related document set. More coherent summaries contain lower percentages of stand-alone concepts, which means that wholist users likely prefer these types of summaries. In addition, analytic users prefer to solve a problem in parts. Therefore, they will prefer a summary that can give equal emphasis for many different parts. Such a summary will produce high entropy values.

C. Procedure

First, we need to compute entropy values and the percentages of stand-alone concepts for all 40 summaries that were used in this paper. Entropy values for all concept nodes of a DG are computed as follows:

$$H(DG) = -\sum_{c} p(c) \log(p(c)).$$

This formula is similar to the entropy computation defined by Navigli and Lapata [29]. In this formula, c is a concept node of a DG, $p(c)$ is computed by the distribution of $\{indeg(c)/2|E|\}$ and, $indeg(c)$ is the in-degree centrality measure that assesses the importance of a concept node c by its connectivity with other concept nodes. This value is defined as the total number of edges that connect to this node. |E| is the total number of edges in a DG.

Second, we use the general linear model with univariate analysis in SPSS as a tool for analyzing these data. In the first analysis, we first choose coherency rating as a dependent variable, cognitive styles as an independent variable, and the percentage of stand-alone nodes as covariates. The result is shown in Table XI. We can see that cognitive styles and the percentages of stand-alone concepts significantly affect the users' coherency ratings (percentage $F = 6.620$, $Sig = 0.010 < 0.05$, and cognitive styles $F = 4.429$, $Sig = 0.012 < 0.05$).

Similarly, in the second analysis, we choose coherency rating as a dependent variable, cognitive styles as an independent variable, and entropy as covariates. The result is shown in Table XII. As we can see, cognitive styles and entropy significantly affect the user's coherency ratings.

TABLE XI
GENERAL LINEAR MODEL ANALYSIS OF THE COGNITIVE-STYLE FACTOR AND PERCENTAGE AS COVARIATES

Descriptive Statistics

Dependent Variable:CoherencyRating

Cognitive Styles	Mean	Std. Dev	N
Wh	3.23	1.268	240
Ana	3.19	1.198	400
Int	3.01	1.072	560
Total	3.12	1.159	1200

Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	20.580[a]	3	6.860	5.159	.002
Intercept	1266.715	1	1266.715	952.633	.000
Percentage	8.802	1	8.802	6.620	.010
CognitiveStyles	11.778	2	5.889	4.429	.012
Error	1590.319	1196	1.330		
Total	13261.000	1200			
Corrected Total	1610.899	1199			

a. R Squared = .013 (Adjusted R Squared = .010)

TABLE XII
GENERAL LINEAR MODEL ANALYSIS OF THE COGNITIVE-STYLE FACTOR AND ENTROPY AS COVARIATES

Tests of Between-Subjects Effects

Dependent Variable:CoherencyRating

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	17.617[a]	3	5.872	4.408	.004
Intercept	1768.539	1	1768.539	1327.557	.000
Entropy	5.839	1	5.839	4.383	.037
CognitiveStyles	11.778	2	5.889	4.420	.012
Error	1593.282	1196	1.332		
Total	13261.000	1200			
Corrected Total	1610.899	1199			

a. R Squared = .011 (Adjusted R Squared = .008)

D. Plan for Integrating the Results to Automatically Determine a User's Cognitive Style

We plan to combine this result with the hidden Markov model (HMM) approach to determine a user's cognitive style [44]. In this approach, HMMs are used to model a user's cognitive style. We create a HMM for wholist users and another HMM for analytic users by training each model using action sequences from representative wholists and analytics in a training set. Then, we identify which cognitive style each user belongs to by comparing how likely his/her action sequence is generated from each of the trained models. One of the potential problems for this approach is that the precision of the categorization heavily depends on the selection of the training set. The training set is manually selected by comparing user's sequences of actions with definitions of analytics and wholists. The manual selection is labor intensive, prone to human errors, and hard to validate. We can address this problem by computing the level of wholist/analytic tendencies in the style of each users using combinations of heuristics from entropy and the percentages of stand-alone nodes, as well as other heuristics such as specificity–generality and height–width factors [44]. The specificity–generality factor of a document indicates how detailed the document is, whereas the height–width factor represents the coverage of topics and the level of detail of a document. In [44], we found that analytics analysts focus on documents with significantly more specific detailed information than wholist analysts. In addition, wholist users start with documents with a wider coverage of topics compared to analytics. We plan to use these heuristics, including entropy and percentages of stand-alone nodes, as indicators of guiding the learning process of the HMMs.

VI. Discussion

With regard to the original hypotheses in this paper, we have found the following results.

1) Informativeness and coherency ratings are significantly different for closely related and loosely related document sets. This result confirms that differences between these two groups of documents sets impact users' judgments about the qualities of the summaries included in the corpus. This result is important in designing any text summarization system that takes the type of document sets into account to enhance a user's satisfaction.
2) Users grouped by the analytic/wholist dimension are significantly different with regard to their assessment of the coherence of the summaries that are generated from different types of document sets. However, users are not much different with respect to the information in these summaries. This result suggests that, although the users in these two user groups generally agree with each other in terms of the main points covered, they have their own preferences in terms of the way that information is presented or written in a summary. In particular, wholist users tend to give significantly better coherency ratings to the summaries that were generated from a closely related document set compared to analytic users. This result may be explained through the follow-up study in this paper. Summaries that are generated from the closely related document set may be more connected than summaries that are generated from a loosely related document set.
3) We found no evidence to suggest that users grouped by the verbal/imagery dimension are significantly different with regard to their assessments of either the way information is presented or the information coverage in different types of document sets. This result suggests that the verbal/imagery dimension may not affect the users' perception on summaries from different types of document sets. One of the possible reasons for this result is that the verbal/imagery dimension has been found to be unreliable [34]. The result may also be caused by the subjectivity of questions to measure this dimension and the lack of a significant contribution of this dimension to individual differences.
4) Graph entropy for concept nodes and the percentages of stand-alone concepts, together with a user's cognitive styles, affect the user's coherency ratings. This result deserves more investigation before a model to determine a user's cognitive styles can automatically be constructed.

In summary, this paper explores the impacts of a user's cognitive styles on assessing a summary, with the hope that the results may be used to improve text summarization techniques. Some ways of adapting to a user's cognitive styles in a text summarization system is desired, because research has shown that a user's cognitive styles affect a user's performance on a web search and during online learning. For example, novice field-dependent users will take longer to find relevant documents compared to novice field-independent users [14]. However, an application that fully adapts to a user's cognitive styles, if it exists, may not improve a user's performance. Some existing research from the human factors community with regard to a user's preferences versus a user's performance (e.g., [2] and [23]) have shown that a user's preferences may not truly reflect his/her performance. What we intend to achieve here is to find a basis for an application that adapts to a user's cognitive styles and improves a user's performance at the same time. Therefore, some tradeoff between adaptivity and performance would be needed.

VII. Conclusion

Cognitive styles have been found to affect a user's information-seeking tasks, reading, and learning processes. In this paper, we have also shown that different users gave significantly different informative ratings and coherency ratings to summaries that were generated from different types of a document sets. This finding has the potential to improve summarization algorithms that do a better job at recognizing key information from collections. More importantly, we found that wholist/analytic users give significantly different coherency ratings to summaries that were generated from the different types of document sets. In addition, the connectivity of a document (measured by graph entropy and the percentage of stand-alone concepts), together with a user's cognitive styles, affected a user's coherency ratings. This result could be used to design a user-centered text summarization system in which we combine both a user's interests and cognitive styles to determine what should be included in a summary.

We are looking to leverage previous user modeling work to build a user model that contains both interests and cognitive styles for a text summarization system. We start with a DG format to capture the most general information in a summary. For document sets, the core information can be constructed through majority vote, whereas for a single document, the core can be created based on the weights of the relation. Summaries that are generated for analytic and wholist users may be generated based on analyzing the links of the resulting DG. Therefore, the summary based on this DG will be biased toward a user's individual interests and a user's cognitive styles. We are currently pursuing this effort and are focused on formally defining the appropriate graph-theoretic measures for expanding DGs from multiple documents.

Appendix
Example of an Abstract With Instruction Given to All Participants

Please read each summary for this document set. For each summary, please rank its informativeness and coherency on a five-point scale.

1) How satisfied are you with this summary in terms of informativeness? (Informativeness refers to how good the

summary can help you answer the specific request given in this topic.)

Most informative				Least informative
1	2	3	4	5

2) How satisfied are you with this summary in terms of coherency? (Coherency refers to how good a summary is structured and organized.)

Most coherent				Least coherent
1	2	3	4	5.

Strong growth in overseas cigarette sales helped RJR Nabisco, the U.S. food and tobacco group, compensate for the continuing effects of lower prices in the domestic market in its first quarter.

The acquisition highlights Philip Morris' strategy of broadening the base of its international tobacco operations, which recently have experienced healthy earnings gains, in contrast to its U.S. tobacco business, where operating income has sharply fallen.

RJR Nabisco, the U.S. tobacco and food group, warned that profits this year from its domestic cigarette business would be hit by the looming cigarette price war in the U.S.

Weiss said he saw little chance that Congress would further restrict advertising by tobacco companies, already proscribed from peddling cigarettes on television and radio and forced to put health warnings on other ads.

This asset is a relic of the time Seita used to have a monopoly of French cigarette distribution, which was abandoned at the end of the 1970s, although it keeps its monopoly on production.

Philip Morris, the U.S. tobacco and food products group, will become the first foreign company to acquire a significant stake in a privatized enterprise in the former Soviet republic of Kazakhstan, under the terms of an agreement announced.

The output of Marlboro in Krakow reached two billion cigarettes, and Philip Morris is offering to invest $200 million in the plant on top of the purchase price.

There must also be doubts about how quickly the government could move toward total prohibition, if it developed the will to do so.

Acknowledgment

The authors would like to thank the reviewers whose suggestions helped improve this paper and Brittany Clark for carefully proofreading this paper.

References

[1] M. R. Amini and P. Gallinari, "The use of unlabeled data to improve supervised learning for text summaries," in *Proc. 25th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, Tampere, Finland, 2002, pp. 105–112.

[2] R. W. Bailey, "Performance vs. preference," in *Proc. Human Factors Ergonom. Soc. 37th Annu. Meeting*, Santa Monica, CA, 1993, pp. 282–286.

[3] P. Brusilovsky and C. Tasso, "Preface to special issue on user modeling for web information retrieval," *User Model. User-Adapted Interact.*, vol. 14, no. 2/3, pp. 147–157, 2004.

[4] J. P. Callan, "Passage-level evidence in document retrieval," in *Proc. 17th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, Dublin, Ireland, 1994, pp. 302–310.

[5] Y. S. Chen and D. R. Macredie, "Cognitive styles and hypermedia navigation: Development of a learning model," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 53, no. 1, pp. 3–15, 2002.

[6] D. A. Cook, "Learning and cognitive styles in web-based learning: Theory, evidence, and application," *Acad. Med.*, vol. 80, no. 3, pp. 266–278, Mar. 2005.

[7] H. Dang, "Overview of DUC 2005," in *Proc. Document Understand. Workshop*, Vancouver, BC, Canada, 2005.

[8] A. Díaz and P. Gervás, "User-model based personalized summarization," *Inf. Process. Manage.*, vol. 43, no. 6, pp. 1715–1734, Nov. 2007.

[9] A. Díaz, P. Gervás, and A. García, "Evaluation of a system for personalized summarization of web contents," in *Proc. 10th Int. Conf. User Model.*, vol. 3538, *LNCS*, 2005, pp. 453–462.

[10] N. Elhadad, "User-sensitive text summarization," Ph.D. dissertation, Columbia Univ., New York, 2005.

[11] A. Field, Discovering Statistics Using SPSS for Windows: Advanced Techniques for Beginners (Introducing Statistical Methods series), Newbury Park, CA: Sage. 2000.

[12] N. Ford, "Learning styles and strategies of postgraduate students," *Brit. J. Educ. Technol.*, vol. 16, no. 1, pp. 65–77, Jan. 1985.

[13] N. Ford, D. Miller, and N. Moss, "The role of individual differences in Internet searching: An empirical study," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 52, no. 12, pp. 1049–1066, 2001.

[14] N. Ford, D. Miller, and N. Moss, "Web search strategies and human individual differences: Cognitive and demographic factors, Internet attitudes, and approaches," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 56, no. 7, pp. 741–775, 2005.

[15] N. Ford, T. D. Wilson, A. Foster, and D. Ellis, "Information seeking and mediated searching—Part 4: Cognitive styles in information seeking," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 53, no. 9, pp. 728–735, 2002.

[16] E. Frias-Martinez, Y. S. Chen, and X. Liu, "Automatic cognitive style identification of digital library users for personalization," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 58, no. 2, pp. 237–251, 2007.

[17] G. Ghinea and S. Y. Chen, "Perceptual multimedia: A cognitive style perspective," in *Digital Multimedia Perception and Design*. Hershey, PA: Idea Group Publ., 2006, pp. 187–205.

[18] G. Ghinea and S. Y. Chen, "Measuring quality of perception in distributed multimedia: Verbalizers vs. imagers," *Comput. Human Behaviors*, vol. 24, no. 4, pp. 1317–1329, Jul. 2008.

[19] M. Graff, "Cognitive style and attitudes towards using online learning and assessment methods," *Electron. J. e-Learning*, vol. 1, no. 1, pp. 21–28, 2003.

[20] M. Graff, "Learning from web-based instructional systems and cognitive style," *Brit. J. Edu. Technol.*, vol. 34, no. 4, pp. 407–418, 2003.

[21] U. Hahn and I. Mani, "The challenges of automatic summarization," *Computer*, vol. 33, no. 11, pp. 29–36, Nov. 2000.

[22] D. Jonassen and L. B. Grabowski, *Handbook of Individual Differences, Learning, and Instruction*. Hillsdale, NJ: Lawrence Erlbaum, 1993.

[23] G. Kissel, "The effect of computer experience on subjective and objective software usability measures," in *Proc. Conf. Companion Human Factors Comput. Syst.*, I. Katz, R. Mack, and L. Marks, Eds., 1995, pp. 284–285.

[24] J. Lee and E. Boling, "Information-conveying approaches and cognitive styles of mental modeling in a hypermedia-based learning environment," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 59, no. 4, pp. 644–661, Feb. 2008.

[25] M. Maňa, M. Buenaga, and J. M. Gómez, "Using and evaluating user-directed summaries to improve information access," in *Proc. 3rd Eur. Conf. Res. Adv. Technol. Digit. Libr.*, *LNCS 1696*, 1999, pp. 198–214.

[26] I. Mani and M. T. Maybury, *Advances in Automatic Text Summarization*. Cambridge, MA: MIT Press, 1999.

[27] H. Martin, "Resource lean and portable automatic text summarization," Ph.D. dissertation, Sch. Comput. Sci. Commun., Stockholm, Sweden, 2007.

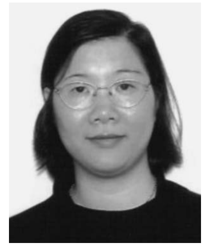
[28] G. A. Miller, "WordNet: A lexical database for English," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, Nov. 1995.

[29] R. Navigli and M. Lapata, "Graph connectivity measures for unsupervised word sense disambiguation," in *Proc. IJCAI*, 2007, pp. 1683–1688.

[30] H. Nguyen, E. Santos, Jr., Q. Zhao, and C. Lee, "Evaluation of effects on retrieval performance for an adaptive user model," in *Adapt. Hypermedia: Workshop Proc. Part I*, Eindhoven, the Netherlands, 2004, pp. 193–202.

[31] H. Nguyen, E. Santos, Jr., Q. Zhao, and H. Wang, "Capturing user intent for information retrieval," in *Proc. 48th Annu. Meeting Human Factors Ergonom. Soc.*, New Orleans, LA, 2004, pp. 371–375.

[32] H. Nguyen, E. Santos, Jr., R. Jacob, and N. Smith, "Evaluation of the effects of user sensitivity on text summarization," in *Proc. IEEE/WIC/ACM Int. Conf.*, Sydney, Australia, 2008, pp. 927–931.
[33] G. Pask, "Styles and strategies of learning," *Brit. J. Edu. Psychol.*, vol. 46, no. 2, pp. 128–148, 1976.
[34] E. R. Peterson, I. J. Deary, and E. J. Austin, "The reliability of Riding's cognitive style analysis test," *Pers. Individual Differ.*, vol. 34, no. 5, pp. 881–891, Apr. 2003.
[35] R. J. Riding and S. Rayner, *Cognitive Styles and Learning Strategies: Understanding Style Differences in Learning and Behaviors*. London, U.K.: David Fulton Publ., 1998.
[36] R. J. Riding and I. Cheema, "Cognitive styles—An overview and integration," *Educ. Psychol.*, vol. 11, no. 3/4, pp. 193–215, 1991.
[37] R. J. Riding and M. Watts, "The effect of cognitive style on the preferred format of instructional materials," *Educ. Psychol.*, vol. 17, no. 1/2, pp. 179–183, Mar.–Jun. 1997.
[38] R. J. Riding and M. Grimley, "Cognitive style, gender and learning from multimedia materials in 11-year-old children," *Brit. J. Educ. Technol.*, vol. 30, no. 1, pp. 43–56, Jan. 1999.
[39] R. J. Riding and S. G. Rayner, *International Perspectives on Individual Differences: Cognitive Styles*. Norwood, NJ: Ablex, 2000.
[40] M. H. Rosa, "Relationships between cognitive styles and reading comprehension of expository text of African American male students," *J. Negro Educ.*, vol. 63, no. 4, pp. 546–555, 1994.
[41] H. Nguyen and E. Santos, Jr., "Modeling users for adaptive information retrieval by capturing user intent," in *Proc. Collaborative Social Inf. Retrieval Access: Techn. Improved User Model.*, M. Chevalier, C. Julien, and C. Soule-Depuy, Eds., 2009, pp. 88–118.
[42] E. Santos, Jr., H. Nguyen, and S. M. Brown, "Kavanah: An active user interface," in *Proc. Intell. Agent Technol. Conf.*, Maebashi, Japan, 2001, pp. 412–423.
[43] E. Santos, Jr., H. Nguyen, Q. Zhao, and E. Pukinskis, "Empirical evaluation of adaptive user modeling in a medical information retrieval application," in *User Modeling*, P. Brusilovsky, A. Corbett, and F. de Rosis, Eds. New York: Springer-Verlag, 2003, pp. 292–296.
[44] E. Santos, Jr., H. Nguyen, F. Yu, D. Li, and J. T. Wilkinson, "Impacts of analysts' cognitive styles on the analytic process," in *Proc. IEEE/ACM Int. Web Intell. Conf.*, Toronto, ON, Canada, 2010, pp. 601–610.
[45] E. Santos, Jr., H. Nguyen, Q. Zhao, and H. Wang, "User modeling for intent prediction in information analysis," in *Proc. 47th Annu. Meeting Human Factors Ergonom. Soc.*, Boulder, CO, 2003, pp. 1034–1038.
[46] E. Santos, Jr., Q. Zhao, H. Nguyen, and H. Wang, "Impacts of user modeling on personalization of information retrieval: An evaluation with human intelligence analysts," in *Proc. 4th Workshop Eval. Adaptive Syst.*, 2005, pp. 27–36.
[47] D. D. Sleator and D. Temperley, "Parsing English with a link grammar," in *Proc. 3rd Int. Workshop Parsing Technol.*, 1993, pp. 277–292.
[48] A. Tombros and M. Sanderson, "Advantages of query-biased summaries in information retrieval," in *Proc. 21st ACM SIGIR Conf.*, 1998, pp. 2–10.
[49] P. Wang, W. B. Hawk, and C. Tenopir, "Users' interaction with World Wide Web Resources: An exploratory study using a holistic approach," *Inf. Process. Manage.*, vol. 36, no. 2, pp. 229–251, Mar. 2000.
[50] H. A. Witkin, C. A. Moore, D. R. Goodenough, and P. W. Cox, "Field-dependent and field-independent cognitive styles and their educational implications," *Rev. Educ. Res.*, vol. 47, no. 1, pp. 1–64, Winter 1977.
[51] F. Wood, N. Ford, D. Miller, G. Sobczyk, and R. Duffin, "Information skills, searching behaviour and cognitive styles for student-centered learning: A computer-assisted learning approach," *J. Inf. Sci.*, vol. 22, no. 2, pp. 79–92, 1996.
[52] Q. Zhao, E. Santos, Jr., H. Nguyen, and A. Mohamed, "What makes a good summary?" in *Comput. Methods Counterterrorism*, S. Argamon and H. Newton, Eds. New York: Springer-Verlag, 2009, pp. 33–50, (Preliminary version presented at HICSS-39 in 2006).



Hien Nguyen received the Ph.D. degree in computer science from the University of Connecticut, Storrs, in 2005.

She is currently an Assistant Professor with the Department of Mathematical and Computer Sciences, University of Wisconsin–Whitewater, Whitewater, where she is actively involved in supervising undergraduate research. Her research interests include user modeling, information retrieval, collaborative information retrieval, recommender systems, intent inferencing, and text summarization, with a current focus on hybrid user models for improving a user's performance in information retrieval.

Dr. Nguyen is a member of the User Modeling-User Adapted Interactions Journal Special Reviewers Board. She is a Member of professional services and committee works, including the Program Committees for the 18th International Conference on User Modeling Personalization and Adaptation (UMAP 2010), 2008–2010 Florida Artificial Intelligence Research Society (FLAIRS) Conference, and the 2006, 2007, and 2011 IEEE International Conference on Systems, Man, and Cybernetics (SMC).

Eugene Santos, Jr. (M'93–SM'04) received the B.S. degree in mathematics and computer science and the M.S. degree in mathematics (specializing in numerical analysis) from Youngstown State University, Youngstown, OH, in 1985 and 1986, respectively, and the M.Sc. and Ph.D. degrees in computer science from Brown University, Providence, RI, in 1988 and 1992, respectively.

He is currently a Professor of engineering with the Thayer School of Engineering, Dartmouth College, Hanover, NH. He is an Associate Editor for the *International Journal of Image and Graphics* and is also on the editorial advisory board for the *System and Information Sciences Notes* and on the editorial boards for the *Journal of Intelligent Information Systems* and the *Journal of Experimental and Theoretical Artificial Intelligence*. His research interest include artificial intelligence, intent inferencing, social and cultural modeling, computational social science, automated reasoning, decision science, adversarial reasoning, user modeling, natural-language processing, probabilistic reasoning, knowledge engineering, verification and validation, protein folding, virtual reality, and active user interfaces.

Dr. Santos is currently the Editor-in-Chief for the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS PART B. He has served on several Program Committees for major conferences on intelligent agents and evolutionary computing.

Jacob Russell received the B.S. degree in management computer systems and mathematics (with emphasis on computers) from the University of Wisconsin–Whitewater, Whitewater, in 2009. He is currently working toward the M.S. degree in computer science in the Department of Computer Sciences, University of Wisconsin–Milwaukee, Milwaukee.

His research interests include artificial intelligence, knowledge engineering, human factors, and user modeling.