ELSEVIER

Personal Publication Assistant: Abstract recommendations by a cognitive model

Action editor: Yiyu Yao

Leendert Van Maanen [*], Hedderik Van Rijn, Maarten van Grootel, Stephanie Kemna, Martin Klomp, Erwin Scholtens

Department of Artificial Intelligence, University of Groningen, P.O. Box 407, NL-9700 AK Groningen, The Netherlands

Abstract

This paper discusses an analysis of how scientists select relevant publications, and an application that can assist scientists in this information selection task. The application, called the Personal Publication Assistant, is based on the assumption that successful information selection is driven by recognizing familiar terms. To adapt itself to a researcher's interests, the system takes into account what words have been used in a particular researcher's abstracts, and when these words have been used. The user model underlying the Personal Publication Assistant is based on a rational analysis of memory, and takes the form of a model of declarative memory as developed for the cognitive architecture ACT-R. We discuss an experiment testing the assumptions of this model and present a user study that validates the implementation of the Personal Publication Assistant. The user study shows that the Personal Publication Assistant can successfully make an initial selection of relevant papers from a large collection of scientific literature.
© 2008 Elsevier B.V. All rights reserved.

Keywords: Recommender systems; User models; Information selection; Human information-processing systems

1. Introduction

In cognitive science, there has been a long tradition to perceive human behavior as a form of information-processing. Within this tradition, human cognitive processes are seen as operating on similar principles or algorithms as computer programs, since both cognition and computer programs have or have been developed to process information. This view has lead to the birth of Artificial Intelligence as an independent research field (McCarthy, Minsky, Rochester, & Shannon, 1955), but has also guided the development of cognitive theories (e.g., Anderson & Milson, 1989; Marr, 1982; Newell, 1990). Even today, the apparent functional overlap between artificial computational systems and the human information-processing sys-

tem is still influential in cognitive theorizing (e.g., Griffiths, Steyvers, & Firl, 2007).

Many cognitive theorists believe that human beings optimize their behavior to successfully cope with the environment (e.g., Anderson, 1990; Marr, 1982; Oaksford & Chater, 1998). This means that, through evolution and learning, human behavior has adapted to be the most suitable behavior in any given circumstance or environment. This is a capacity also desirable in artificial systems design, especially when these systems have to operate on an unknown or dynamic environment. Therefore, computer scientists and artificial intelligence researchers have studied how computer systems can optimize their behavior as well (e.g., Goldberg & Holland, 1988; Kohonen, 2001).

A domain that has not benefited that much from this cross-fertilization is the problem of selecting relevant information, either for oneself or for others. The research field that studies how to disclose relevant information is known as Information Retrieval (Salton & McGill, 1983). A

---
[*] Corresponding author.
E-mail address: leendert@ai.rug.nl (L. Van Maanen).

typical field in which the problem of selecting relevant information arises is the scientific community. For example, the number of scientific publications in the relatively small ISI subject category *Information Science & Library Science* was 2054 in 2006.[1] This means that researchers working in this area have to read (or at least scan through) over two thousand papers a year to keep up with the current developments. However, this number is, if anything, an underestimation of the total number of potentially relevant papers, as this number only holds if the researcher is interested in a single subject area. In practice, most researchers work on the intersection of multiple domains, increasing the number of potentially relevant papers enormously. In general, because of the continuous increase of storage capacity for digital media, and the increased availability of digital or digitized media sources, companies, institutions, and individual people are being confronted with an increase in the amount of information that potentially is relevant to their purposes.

In this paper, we will describe a system that partly solves this problem for the scientific domain: Our system selects relevant scientific papers from a large collection of scientific abstracts. Instead of working from a pure computer science perspective, we will present a system that is based on constraints from cognitive theories. In particular, we chose to follow the rational analysis approach (Anderson, 1990; Oaksford & Chater, 1998), as incorporated in the ACT-R architecture of cognition (Anderson, 2007a). The rational analysis approach states that human memory is optimally adapted to fit the needs of the environment we live in, based on the interactions of the cognitive agent with the environment in the past. This approach has been successfully applied to predict various aspects of human behavior [e.g., as reviewed by Chater and Oaksford (1999)].

We will begin with an analysis of how users behave when engaging in the selection of information. Next, we will discuss how the ACT-R cognitive architecture incorporates rational analysis, and how this can be applied to information selection. We will continue with an outline of an application based on the resulting model, the Personal Publication Assistant (or Publication PA for short), and how this application behaves under different conditions, as well as a user study that will demonstrate the applicability of our approach in a real world setting. In the last section, we will discuss in what way the Publication PA deviates from other approaches towards the task of matching papers to researchers, or vice versa.

1.1. Information selection

An example of the problem addressed in this paper is the selection of relevant information when attending a large, multi-track scientific conference. Often, an attendee finds him or herself overwhelmed by the amount of presenta-

tions that can be attended. With so little time to find the talks that are really interesting, chances are that one ends up in the wrong track, listening to presentations that hardly kindle ones interest, while in another track relevant work is being discussed. Although this might bring unforeseen beauty, often a better selection of relevant work would be preferable. There are solutions to this problem, such as giving the attendees the proceedings well in advance so they have more preparation time. However, this solution is often not viable due to practical constraints. A better solution might be to provide an automatic recommendation based on the personal interests of the conference attendees, which is the approach that will be discussed in this paper.

To build a successful recommendation system, it is important to know how the selection process takes place in unsupported settings. The information selection process starts when a researcher registers at a conference and receives a copy of the conference proceedings. Based on informal analyses, the next step is to perform a quick scan of all titles, author names, or abstracts for words or names that are familiar. If an entry contains enough interesting words, it is selected for further and more careful reading. Obviously, the assumption that is made implicitly is that individual words in the abstract accurately reflect the contents of the paper or presentation. Ries et al. (2001) have shown that this assumption holds for abstracts and papers, at least in the medical domain. In order to determine if a word qualifies as interesting in the context of the conference, the researcher might assess whether she has used the word in her own research in the past. One could say that the researcher tries to discover the degree of familiarity she has with an abstract, and if that degree of familiarity is high enough, she selects that presentation as potentially worthwhile to visit.

To assist a researcher in the information selection task, we propose a model of the recognition aspects of the task. That is, we propose a model that makes a preselection from the available information based on a notion of familiarity adapted to the individual researcher. To achieve this, we will develop models of the declarative memory systems of individual researchers (henceforth referred to as user models) and of the process of recognizing words. Each user model can be seen as a representation of an individual researcher's interests, as it incorporates the frequency, recency, and context of the words used by the researcher to describe her research. In previous research (Anderson & Milson, 1989; Anderson & Schooler, 1991), a formal model has been developed of how the retrieval of declarative facts from memory can be described. In the next section, we will give a detailed overview of that model, but we will highlight the two most important aspects here. One key idea is that declarative memory is optimally adapted to serve the needs of the cognitive agent (Anderson, 1990; Oaksford & Chater, 1998). The other is that most facts in declarative memory are initially formed by perception (Anderson & Schooler, 1991). Combined, this means that the adaptive nature of declarative memory is

---

[1] *Source:* ISI Web of Knowledge, retrieved 19-12-2007.

essentially a reflection of the perceptions of the cognitive agent. As a consequence, this means that looking for structure in the environment can derive the structure of declarative memory.

1.2. Rational analysis of memory

Anderson and Schooler (1991) showed that the probability that a memory will be needed in the near future depends on the pattern of prior exposures to the piece of information stored by that memory. For example, the probability that someone will contact you by email today depends on the frequency and recency of her emails to you in the past (Anderson & Schooler, 1991). Likewise, the probability that you will need some declarative fact from memory right now depends on the frequency and recency of the prior usage of that fact. Both relations are captured by Eq. (1), in which B stands for the base-level activation (reflecting the probability), t_i stands for the time since exposure to event i, and d represents the speed with which the influence of each exposure decays. The summation is over all (n) previous encounters of the events (i):

$$B = \ln \left[\sum_{i=1}^{n} t_i^{-d} \right] \tag{1}$$

Besides frequency and recency of usage of declarative facts, the context in which these facts occur also plays a role in the activation of these facts. This activation component will be called the spreading activation (Quillian, 1968), and represents the likelihood that one declarative fact will be needed if another one is currently being used. These likelihoods depend on the pattern of prior exposures with the declarative facts, as represented by the relatedness measure R_{ji} between two facts j and i (Anderson & Lebiere, 1998; Anderson & Milson, 1989):

$$R_{ji} = \frac{F(W_j \,\&\, W_i) F(N)}{F(W_j) F(W_i)} \tag{2}$$

where $F(W_j)$ and $F(W_i)$ are the frequencies of respectively fact j and i, $F(N)$ the total number of exposures and finally $F(W_j \,\&\, W_i)$ is the number of co-occurrences of the facts j and i. Eq. (2) is sometimes referred to as associative strength (Anderson & Lebiere, 1998; Anderson & Milson, 1989), to indicate that the relatedness between two facts is determined by the environment. The model of declarative memory outlined here has been successfully deployed in predicting behavior in a variety of memory-related cognitive tasks (e.g., Anderson, Bothell, Lebiere, & Matessa, 1998; Anderson & Schooler, 1991; Van Rijn & Anderson, 2003).

2. Implementation of the Personal Publication Assistant

The Personal Publication Assistant is a personalization tool based on a personalized rational analysis of memory.

Therefore, the user models underlying the recommendations are constructed on an individual basis. In these models, each word that occurs in one of the abstracts of the user is represented by a combination of base-level activation (adapted from Petrov, 2006) and spreading activation from the other words in the model (Anderson & Lebiere, 1998). These activation values can be calculated using the statistical properties of the words in the published abstracts of an individual researcher:

- The year in which it appears for the first time in one of the user's abstracts.
- The year in which it most recently appears in one of the user's abstracts.
- The frequency of appearance.
- The frequency of co-occurrence with another word.

Based on these properties, we create an individual representation of a researcher's interests using the rational analysis described above. The Publication PA applies these individual user models to predict the relevance of words that occur in other scientific abstracts, by calculating how familiar these abstracts are. In the next sections, we will describe in more detail how the Publication PA calculates the base-level and spreading activation values, which words from the abstracts are taken into consideration, and how the system comes to a selection of the relevant information.

2.1. The relevance of individual words in the user model

With the equations that are provided by the rational analysis approach to declarative memory, we can calculate the base-level activation of a word based on its occurrences in publications of the user. The base-level activation can be seen as a measure of interest, with the most interesting words having the highest base-level activation.

For this application, an optimized version (Petrov, 2006) of the base-level equation discussed earlier (Eq. (1)) was used. In this equation (Eq. (3)), the decay parameter is fixed at 0.5 (and is reflected in Eq. (3) as the square root operators) and a history parameter (h) is added:

$$B = \ln \left[\frac{1}{\sqrt{t_1 + h}} + \frac{2n - 2}{\sqrt{t_n} + \sqrt{t_1 + h}} \right] \text{ with } h > 0 \tag{3}$$

The first component of this equation reflects the most recent encounter of that word: The longer ago the word was encountered, the smaller the contribution is. The second component reflects the frequency of usage of the word. This optimized version of the base-level activation equation assumes that the encounters of the word are evenly spaced over time between the first encounter and the last encounter of the word. In the default equation, the base-level activation is a product of both recency and frequency. However, in a recommendation system, it might be useful to be able to change the balance between both factors. For example, a researcher might still be interested in work

relating to older work, even though a recent project has resulted in a set of papers on a new topic. To enable this, we added the history parameter. The history parameter influences the effect of recency. Informally, a higher value for *h* spreads the publications over a longer time frame, decreasing the relative activation of a word that only recently came up in analyzed texts. In Experiment 1, we will demonstrate that the *h* parameter is an important parameter when recommending papers with the Publication PA.

2.2. The influence of context on word relevance

Apart from the frequency and recency of usage of a word, the context in which a word occurs is also important. For instance, using the word *model* in your paper on user models should not elicit conference talks on fashion models. So, context words – like in this example *user* or *rational* – are important in determining the activation of words such as *model* or *analysis*. The context in which a word has occurred in previous abstracts is incorporated in the model by spreading activation (Eq. (2)), which reflects the personalized probability that a word will be needed in connection with another word.

Recommendations occur by mediating the base-level activation of a word with the spreading activation of other words:

$$A_i = B_i + \sum_j WR_{ji} \tag{4}$$

In Eq. (4), the base-level activation of the word *i* in a specific abstract is increased with the sum of all weighted connections with the words also found in that abstract. The connections are weighted because otherwise the ratio between the base-level activation and the spreading activation would be dependent on the number of associations. For this application the base-level activation of the connecting word (*j*) is used as the weight (*W*), to scale down with the spreading activation from words that have a low base-level activation. This would be the case when the word *i* co-occurred often in the past with a word *j* that is present in the current abstract but which is not often used anymore (i.e., has a low base-level activation). This would cause the spreading activation to be high while the connection is less relevant at the current time, negatively influencing the selection of relevant papers.

2.3. Filtering of non-content words

The relatedness measure R_{ji} has shown to be a robust method of boosting the base-level activation as a function of the connectedness. That is, if two words always occur in tandem, the activation of the second word will be boosted when the first word is encountered. At the same time, a word that occurs in combination with many other words spreads less activation. In normal word usage, words as *the* and *is* spread only a small amount of activation because

of this. This effect makes sure non-content words do not influence base-level activations of other words too much. However, a problem might arise when the formulation of sentences in scientific abstracts differs from normal word usage. Because of spatial constraints, word usage in scientific abstracts might differ from normal written English. This might result in a lower frequency of function words, increasing their spreading activation (Eq. (2)), with a possibly negative influence on the eventual recommendations. To counter the unwanted influence of normally high-frequent words, these words are filtered from the data using a lexical database (Baayen, Piepenbrock, & Van Rijn, 1993). An analysis of the frequency distribution of words in both scientific abstracts and normal written English will demonstrate that filtering out high-frequent words will not interfere with how well an abstract represents the contents of a paper.

2.3.1. Analysis

To compare word usage in scientific abstracts with word usage in normal lexical content, the abstracts of all publications that appeared in the *Cognitive Science Journal* between 2004 and 2006 were used. Numeric symbols and punctuation were removed from the abstracts, resulting in a list of the words that were used in the abstracts. For each word, the frequency in all the abstracts was contrasted with an estimate of the normal frequency in written English, taken from the CELEX lexical database (Baayen et al., 1993). If a word was not found in the database because of spelling mistakes or terminology, the CELEX frequency was assumed 0, and the frequencies of multiple occurrences of a word were summed because in CELEX the frequencies of homonyms are counted separately. The CELEX frequencies were scaled to the total number of words of the abstracts to make them comparable.

2.3.2. Results

In Fig. 1, the ratio between the CELEX word frequencies and the abstract word frequencies is plotted. We used a logarithmic scale for easier presentation. Fig. 1 visualizes that the usage of words in scientific abstracts differs from the distribution of words used in normal written text. If the distributions were similar, then the dashed horizontal line would have represented the ratio. However, it becomes clear that a large part of the words used in the abstracts occur less often in normal written English; those are the words with a frequency ratio below one. Only a small part of the words occurs more often in normal written English.

Thus, 2190 of the words used in the abstracts of the *Cognitive Science Journal* between 2004 and 2006 occur more frequently in scientific abstracts than in normal written English, while only 412 words occur more often in normal written English. However, those 412 words account for a large portion of the total amount of word occurrences found in the CELEX database (440,000 of the total of 740,000 occurrences of these words), while the 2190 words that are less frequent in normal written English generate
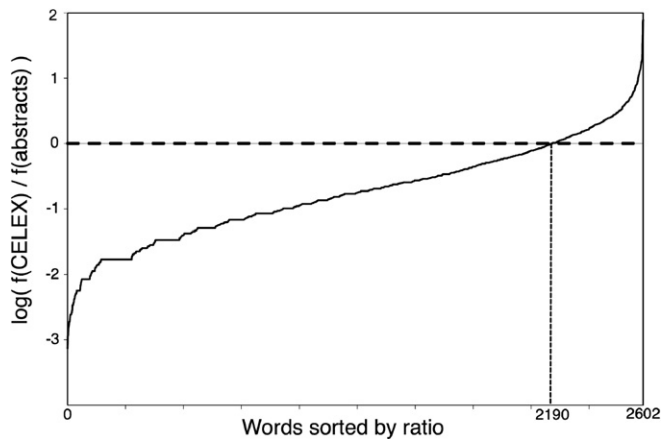
Fig. 1. Log ratio of word usage frequencies in scientific abstracts and normal written English (CELEX), sorted by increasing frequency ratio. The dashed line indicates when words are used more often in scientific abstracts than in normal usage.

less occurrences than the 412 high-frequent words (300,000 of 740,000 word occurrences). This difference is caused by abstracts containing jargon and the tendency to use as little function words as possible, whereas in normal language these words are used very frequently. Thus, removing the words from the scientific abstracts that are most frequent in normal written English will not remove any of the important content words, as only words above the dashed line in Fig. 1 are deleted, while words below the dashed line in Fig. 1 are the words that are relevant to the Publication PA.

2.4. Selection of relevant abstracts

The final part in the recommendation is finding the amount of activation for each paper and presenting the user with a ranking or selection. In general, abstracts in which many words have a high activation, have a high degree of familiarity to the researcher, and are thus interesting enough to select. To compare the relevance of papers with each other, every abstract has to be represented by a single value. One solution would be to sum the activations of all the words in the conference abstract. However, simply summing activation values would result in a bias towards longer abstracts. To counteract this bias, we chose to average the activation of the words that occur in both the abstract and the user model. This means that the effect of abstract length is neutralized, while still taking all activation values of the words in the abstracts into account.

3. Experiments

To validate the Publication PA, we first analyzed what the influence of the *h* parameter is. Second, we performed a user study with a sample of researchers from the field of cognitive science, asking them to rate how much a recommended abstract aligned with their interests.

3.1. Experiment 1: History parameter analysis

3.1.1. Methods

We analyzed the behavior of the Publication PA with four different values for the history parameter: $h = 0.0001, 0.1, 10$ and 1000. The parameter values were chosen to maximize a potential effect. The only other parameter in the system (the decay parameter d) was left at the default value of 0.5.

As a test set, we took the abstracts of the publications of Professor John R. Anderson, for as far as indexed by PsycINFO.[2] When visually inspecting his publication record, it shows some stable interests over time, but also some changes in interest. As a cognitive modeler, almost all of Anderson's publications deal with cognition and the cognitive architecture he developed, ACT-R. However, a change in focus can be observed. From the start of his career, Anderson's interests seem to be related to learning and memory (as witnessed by for instance Anderson & Bower, 1972, 1973), whereas more recently he seems to have developed an interest in functional brain imaging techniques (for instance, Anderson, 2007b; Anderson, Albert, & Fincham, 2005). These trends should also be visible if we apply different parameter values to the *h* parameter and construct different user models.

Since for this analysis we are interested in the contribution of different words to a researcher's interests, we only computed the activation values of each word in the user models. By contrast, the Publication PA also averages over these activation values insofar they occur in an abstract, to get an estimate of the relevance of that abstract. This aspect is included in the user study presented in Experiment 2.

3.1.2. Results

To compare the user models that were constructed with the various values for the *h* parameter, we ordered the words in the user models according to their activation values. Thus, the ordering represented the estimated importance of a word for a person's interest. Fig. 2 presents the rank order values of various words that are exemplary of the trends found in Anderson's publication record. Small values of *h* indicate that the relative influence of more recent publications increases; this effect is reflected by the decreasing rank (and thus increasing importance) of the words *functional* and *imaging* for decreasing values of *h*. These words do all relate to the recent research interests. On the other hand, the words *memory* and *experiments* show the opposite trend. This reflects a shift of interest from prototypical memory-related research in which multiple experiments are presented per paper. Also, the ranks of some words stay constant with changing *h* values. *ACT-R* and *cognitive* are words that appear in both recent and past abstracts of Anderson, indicating a stable interest in these concepts.
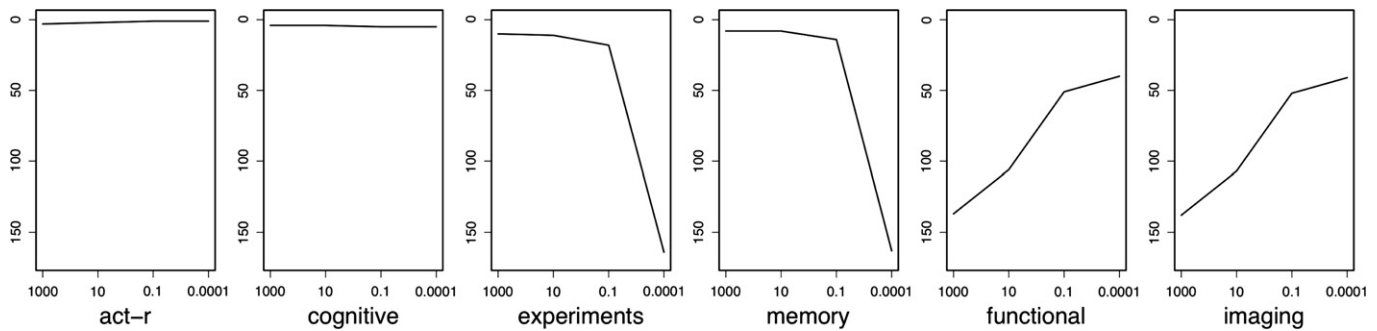
---

[2] http://psycinfo.apa.org/.

Fig. 2. Rank order of the activation values of example words that occur in the user models created for one researcher with varying history parameter values. The history parameter values we tested are indicated on the *x*-axis. Low rank order values indicate that a word is important for determining the researchers interests. If *h* decreases, the relative influence of words that were used in the past also decreases, and the influence of recently used words increases.

This qualitative inspection of the results leads us to believe that the history parameter plays an important role in the selection of relevant abstracts, because it determines the ranking of the activation values. What the optimal setting for this parameter should be might be determined in a large user study in which we ask participants to rate the relevance of abstracts that are selected using various values for the history parameter (as has been done for this analysis). However, given the personal nature of interest, it seems better to leave the optimal setting to the user, for example, by presenting the user with the possibility to set this parameter in the user interface. To evaluate the performance of the Publication PA independent of the relative importance of word usage history, we decided to run the user study with *h* set to 10.

3.2. Experiment 2: User study

We performed a user study to evaluate the recommendations provided by our abstract recommender system. We asked 10 researchers (2 full professors, 2 associate professors, 5 assistant professors, and 1 post-doc) from various subfields of cognitive science and from various countries how much they were interested in a paper after reading the abstract.

3.2.1. Methods

For each of the researchers, we constructed user models based on the abstracts of their published work insofar it was indexed by PsycINFO. Next, we ordered all abstracts from the last three volumes (2004–2006) of the *Cognitive Science Journal* according to their relevance for an individual researcher, based on the researcher's published abstracts.

From the ordered list of abstracts, we presented the top five abstracts, the bottom five abstracts (that is, the least relevant abstracts), and five abstracts from the middle of the list to each researcher. The presentation order of these 15 abstracts was randomized, to eliminate any effects from expectations about the presentation order. We asked the researchers to indicate with a grade between 0 and 9 how much they were interested in the papers, based on the

abstracts. We adopted this scale from similar work done by Dumais and Nielsen (1992) in order to be able to make a comparison between their approach and ours. Following Dumais and Nielsen (1992, p. 235), we characterized the meaning of the rates as follows:

- 8–9: right up my alley
- 6–7: good match
- 4–5: somewhat relevant
- 2–3: I'm following it, sort of
- 0–1: how did I get this one?

3.2.2. Results

To analyze the performance of the Publication PA, we applied two measures of relevance:

- mean rated relevance
- precision

The precision and mean rated relevance were applied to each of the three groups (top 5, middle 5, bottom 5). Because it is not feasible for the participants to rate all available abstracts from the *Cognitive Science Journal* between 2004 and 2006 (129 abstracts), we did not calculate the rate of recall, as is often used in these kinds of applications (Salton & McGill, 1983). However, the recall rate is implicitly accounted for in the measures we did apply.

3.2.2.1. Mean rated relevance. We analyzed the relevance rates given to the abstracts for each group. Fig. 3 shows the means of the rates per group. Welch *t*-tests between the groups reveal that the rates given for the top 5 abstracts differ significantly from the other two groups ($t = 4.20$, d.f. = 86.54, $p < 0.001$ for the top 5 vs. the bottom 5 and $t = 3.64$, d.f. = 94.06, $p < 0.001$ for the top 5 vs. the middle 5). The rates for the bottom five abstracts did not differ significantly from the rates for the middle five abstracts. This is in line with the observation that in multidisciplinary journals such as Cognitive Science, the relevance rate does not decrease linearly, but instead that only a small part of the published papers is relevant for a researcher, and the rest is not.
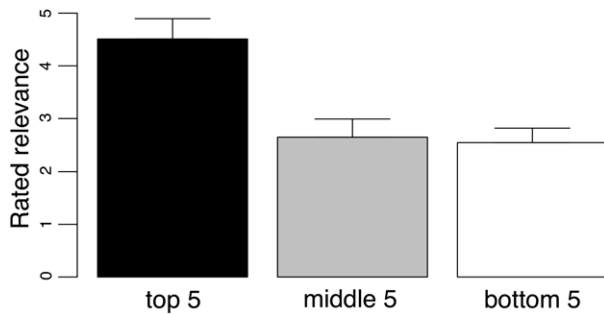
Fig. 3. Mean ratings per group. Error bars denote standard errors. The mean ratings that the participants provided for the top 5 recommended abstracts is significantly higher than for the other two groups.
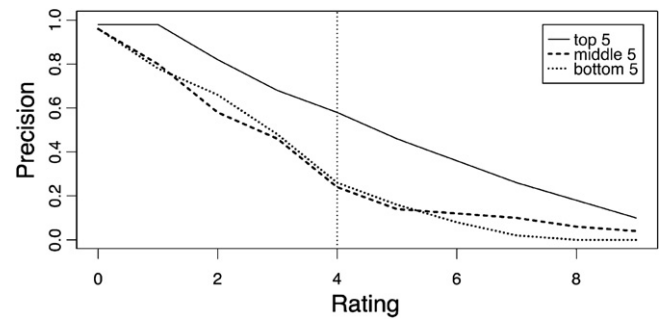

Fig. 4. Precision of the Publication PA for different interpretations of relevance. The dotted vertical line indicates the point at which relevance is interpreted as *somewhat relevant* or better (rating 4 on the scale provided to the participants). The figure shows that for all interpretations of relevance, the precision for the top 5 recommended papers is higher than for the other two groups.

If the Publication PA would not be able to suggest relevant papers, this would mean that in all three groups the number of highly rated papers would be equal on average. However, if this were the case, we would not be able to observe significant differences in the mean rated relevancies between the top 5 recommended papers and the other two groups. The fact that we did find this difference indicates that the system is able to provide a meaningful rank order in which the higher rated papers will be ranked higher.

3.2.2.2. Precision. Precision of retrieval is usually defined as the number of relevant documents that is retrieved relative to the total number of documents retrieved (Salton & McGill, 1983). Following the meanings of the anchor points of the scale we provided to the participants, relevance should be taken as rated with 4 or higher. Using Eq. (5), the precision of the Publication PA in the top 5 recommended abstracts is 0.58. Welch *t*-tests show that this differs from the precision of the middle 5 recommended abstracts (precision = 0.24, $t = 2.94$, d.f. = 17.63, $p = 0.009$) as well as the bottom 5 recommended abstracts (precision = 0.26, $t = 2.35$, d.f. = 17.45, $p = 0.03$):

$$P = \frac{|\{rating > X\} \cap \{retrieved\ abstracts\}|}{|\{retrieved\ abstracts\}|} \qquad (5)$$

Because this notion of relevance may be considered arbitrary, we also calculated the precision of the Publication PA with different assumptions on relevance. For example, we calculated precision under the assumption that only abstracts rated 8 or higher were relevant, or that all abstracts rated 2 or higher were relevant. In Fig. 4, the results of this analysis are presented. The figure shows that, although precision declines with a more stringent notion of relevance, the precision in the top 5 recommended abstracts is always higher than in the other two groups ($\chi^2(18) = 35.66$, $p = 0.008$).

4. Discussion

With our experiments, we demonstrated both the flexibility of the Publication PA and its applicability. With only one parameter, we could change the recommendations of the system in such a way that the relative influence of older papers changed, resulting in different recommendations.

With the *h* parameter at a fixed value, we demonstrated that the Publication PA can provide meaningful recommendations for individual users. Two observations from this experiment should be further discussed.

From both the precision measure and the mean ratings, it becomes clear that there is no real difference between the group of abstracts from the bottom of the order list of abstracts from *Cognitive Science Journal* (2004–2006) and the 'middle' group. This shows that from a large collection of papers, only a very small subset is relevant for a particular user, underlining the need for filtering mechanisms or recommender systems.

Fig. 3 shows that the mean rated relevance for the top 5 recommended abstracts is 4.5. This qualifies as *somewhat relevant*, but not *right up my alley*. We contribute this to the nature of the data set we used to recommend abstracts from. The *Cognitive Science Journal* is a highly multidisciplinary journal, accepting papers from a wide range of research areas (as witnessed for instance by the set of keywords authors can use when submitting, published on the website of the Cognitive Science Society).[3] As a result, papers addressing very specific topics, that may be *right up my alley*, will be presented to other, more specialized, journals. Thus, the ratings provided by our participants might be a bit lower than expected, because abstracts that would be rated as *right up my alley* were probably underrepresented in the data set.

4.1. Related work

The problem of matching researchers and papers has been addressed before, in the context of systems that use Latent Semantic Indexing (LSI) (Dumais, 2003; Dumais & Nielsen, 1992; Foltz & Dumais, 1992). Our approach deviates from these earlier attempts in a number of ways. LSI assumes that the similarity of two documents is

---

[3] http://www.cognitivesciencesociety.org/keywords.html.

reflected by the similar word frequency distributions that are manifest in these documents (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990; Landauer & Dumais, 1997; Landauer, Foltz, & Laham, 1998). However, instead of taking the raw frequency statistics into account, LSI performs a mathematical analysis (singular value decomposition) that is capable of higher-order inference. That is, LSI calculates the probability of each of the words occurring in a document, given multiple documents.

Instead of LSI, the measure of semantic relatedness that we apply, associative strength (Anderson & Lebiere, 1998; Anderson & Milson, 1989), is equivalent to Point-wise Mutual Information (PMI), given a reasonably large data set (Farahat, Pirolli, & Markova, 2004). PMI is also based on the statistical properties of the documents, but, in contrast to LSI, PMI is a direct measure of the likelihood that one word will occur, given the presence of another. As a measure of *semantic similarity*, PMI has been shown to perform equal to or better than LSI (Turney, 2001). We expect therefore that PMI will also be a better representation of semantic relatedness than LSI (cf. Van Maanen, Borst, Janssen, & Van Rijn, 2006).

Besides the method of calculating the semantic relatedness, also the corpus of text on which it is performed differs. Dumais and Nielsen (1992) and Foltz and Dumais (1992) used a fixed semantic space for all users of their system. Recently, however, it has been shown that the choice of corpus greatly influences the semantic distance, even when applying the same measure of semantic relatedness (Lindsey, Veksler, Grintsvayg, & Gray, 2007). By contrast, we constructed personalized semantic spaces for individual users. That is, the associations between words in the semantic space reflect the semantic relatedness as apparent from the statistical properties of word usage in the abstracts of one user. This obviously will result in more individualized recommendations, because only the associations between words that a single researcher would also make, are present. Also, the problem of corpus selection does not arise, because the corpus used is already the best possible representation of a researcher's interest, namely her own publication record.

When it comes to the performance of the Publication PA as compared to the approach taken by Dumais and Nielsen (1992), the Publication PA seems to perform equally well. Dumais and Nielsen (1992) report a precision of 0.51, slightly lower than our value of 0.58. However, the computation of precision differs between the two approaches. In general, comparison is difficult because of the different nature of the data sets used. While the abstracts from the *Cognitive Science Journal* are very multi-disciplinary and thus very diverse, the abstracts used by Dumais and Nielsen (1992) are from a very specialized conference (*ACM Hypertext'91*). This difference in diversity of topics included in the data sets could explain the difference in mean rated relevance between the Publication PA (4.5) and the system by Dumais and Nielsen (5.75). In the Dumais and Nielsen experiment, both the abstracts and the researchers they

are being assigned to, are specialized in hypertext. Therefore, the mean relevance of the data set for the researchers is already higher than in our experiment.

To a certain extent, our work bears resemblance to the work of Pirolli and colleagues towards Information Foraging (Fu & Pirolli, 2007; Pirolli, 2005; Pirolli & Card, 1999; Pirolli & Fu, 2003). They provided a rational analysis of how users search for relevant information, and applied this to information search on the World Wide Web. This way, they were able to model web navigation aspects of a typical user. In Information Foraging theory, the likelihood that a certain document or webpage is relevant is based on the base-level activation of the words in that document and the spreading activation from the words in that document to the words in the search query. Similarly, the Publication PA computes the relevance of a paper based on the base-level activation of the words in the abstract and the spreading activation from the words in that abstract to the words in the user model. One of the important components of the Publication PA is the construction of the user model, which ensures that only words that are relevant for an individual researcher are considered in computing the relevance of an abstract.

However, the Information Foraging models differ in that they capture the information search behavior of a typical human being serving the web, whereas the Publication PA is a personalization tool, and is intended to model the information needs of an individual researcher. As outlined above, the semantic relatedness estimates applied by the Publication PA are therefore personalized for each individual researcher, resulting in different behavior of the model for each researcher.

5. Conclusion

In this paper, we proposed a method for the personalization of information selection, based on rational analysis and cognitive architectures. We developed an application, the Personal Publication Assistant (Publication PA), for the recommendation of relevant scientific abstracts to researchers, based on their publication record to date. In two experiments, we analyzed the behavior of the Publication PA and found that it is a flexible and adaptable system, as well as an adaptive system. From Experiment 1 we concluded that users of the Publication PA can adapt the nature of the recommendations to their own personal wishes, using only one parameter. In a final version of the interface, this parameter could be controlled by a slider bar. Some researchers might be only interested in their current topic, for instance because they have just switched research topics. They can choose a low value for this *h* parameter. Researchers that would rather want to follow what is being published in research fields they previously published in may choose a high value of the *h* parameter.

Experiment 2 demonstrated that the Publication PA can select relevant papers for individual researchers. Papers that were recommended by the system were rated higher

by the participants than papers that were not recommended.

The techniques applied in the Publication PA might also be applied to develop recommender systems in other domains in which personalized information retrieval is desirable. The domain should be primarily characterized by textual information sources, such as conference or journal papers, and the users should also be characterized by textual testimonials of their interests. Two examples of the wider applicability of the method of information selection that we proposed here are the problem of assigning manuscripts submitted to a conference to reviewers, and the problem of selecting relevant press bulletins from the stream of bulletins provided by press agencies world wide. We will discuss both these examples and hint at an implementation of our technique.

The assignment of manuscripts submitted to a conference to reviewers is a problem very similar to the selection of relevant abstracts for a reviewer. Even though reviewers can often indicate their areas of expertise, it is hard for conference program chairs to match every submission to the most qualified reviewers. Since the area of expertise of a reviewer is reflected in his or her publication record, user profiles that reflect the areas of expertise could be generated based on the publication record. By matching the profiles against each submitted abstract, the best-suited reviewer for each abstract will be associated with the highest relevance score. This way, conference chairs can easily assign submitted manuscripts to reviewers without having to rely on the reviewer's own opinion of his or her expertise, or without having to burden them with long questionnaires about their fields of research.

Press agencies produce many bulletins a day, often over 12,500 bulletins a year.[4] A reporter trying to read the most important press bulletins for his or her interests has to make a selection from this vast amount of information. Although press agencies often tag their bulletins or assign them to a certain category, it is easy to miss the one that is important. By creating profiles of reporters based on the news articles they have written over the years, an application similar to the Publication PA could make a meaningful selection for them.

A cognitive model of information selection can thus guide the development of a recommender system, because it provides insights in which features from the pieces of information are relevant for the selection process. The analysis suggests that the features that people that are engaged in retrieving relevant information use are the history of usage of words, and the co-occurrence of words. By incorporating these features in the same way as a cognitive model of human memory does, we have created a successful Publication PA, that for example can decrease the work load of individual researchers attending a conference by creating a preselection in the conference proceedings.

---

[4] *Source:* ANP Press support.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (2007a). *How can the human mind occur in the physical universe?* New York: Oxford University Press.

Anderson, J. R. (2007b). Using brain imaging to guide the development of a cognitive architecture. In W. D. Gray (Ed.), *Integrated models of cognitive systems* (pp. 49–62). New York: Oxford University Press.

Anderson, J. R., Albert, M. V., & Fincham, J. M. (2005). Tracing problem solving in real time: fMRI analysis of the subject-paced tower of Hanoi. *Journal of Cognitive Neuroscience, 17*(8), 1261–1274.

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language, 38*(4), 341–380.

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free-recall. *Psychological Review, 79*(2), 97–123.

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Washington: Winston and Sons.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah: Lawrence Erlbaum.

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review, 96*(4), 703–719.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2*(6), 396–408.

Baayen, H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical database* (CD-ROM).

Chater, N., & Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends in Cognitive Sciences, 3*(2), 57–65.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41*(6), 391–407.

Dumais, S. T. (2003). Data-driven approaches to information access. *Cognitive Science, 27*(3), 491–524.

Dumais, S. T., & Nielsen, J. (1992). Automating the assignment of submitted manuscripts to reviewers. In *Proceedings of the 15th annual international SIGIR*. Denmark: ACM.

Farahat, A., Pirolli, P., & Markova, P. (2004). *Incremental methods for computing word pair similarity*. Technical report no. TR-04-6, Palo Alto Research Center, Palo Alto, CA.

Foltz, P. W., & Dumais, S. T. (1992). Personalized information delivery: An analysis of information filtering methods. *Communications of the ACM, 35*(12), 51–60.

Fu, W. T., & Pirolli, P. (2007). SNIF-ACT: A cognitive model of user navigation on the World Wide Web. *Human–Computer Interaction, 22*(4), 355–412.

Goldberg, D. E., & Holland, J. H. (1988). Genetic algorithms and machine learning. *Machine Learning, 3*(2), 95–99.

Griffiths, T. L., Steyvers, M., & Firl, A. (2007). Google and the mind: Predicting fluency with PageRank. *Psychological Science, 18*(12), 1069–1076.

Kohonen, T. (2001). *Self-organizing maps*. Berlin: Springer.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211–240.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*(2–3), 259–284.

Lindsey, R., Veksler, V. D., Grintsvayg, A., & Gray, W. D. (2007). Be wary of what your computer reads: The effects of corpus selection on

measuring semantic relatedness. In *Proceedings of the eighth international conference on cognitive modeling*.

Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman.

McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. E. (1955). *A proposal for the Dartmouth summer research project on artificial intelligence*.

Newell, A. (1990). *Unified theories of cognition*. Cambridge: Harvard University Press.

Oaksford, M., & Chater, N. (1998). *Rational models of cognition*. Oxford: Oxford University Press.

Petrov, A. A. (2006). Computationally efficient approximation of the base-level learning equation in ACT-R. In D. Fum, F. Del Missier, & A. Stocco (Eds.), *Proceedings of the seventh international conference on cognitive modeling* (pp. 391–392).

Pirolli, P. (2005). Rational analyses of information foraging on the web. *Cognitive Science, 29*(3), 343–373.

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review, 106*(4), 643–675.

Pirolli, P., & Fu, W. T. (2003). SNIF-ACT: A model of information foraging on the World Wide Web. In *Proceedings of the ninth international conference on user modeling*.

Quillian, M. R. (1968). Semantic memory. In M. Minsky (Ed.), *Semantic information processing* (pp. 216–270). Cambridge: MIT Press.

Ries, J. E., Su, K., Peterson, G., Sievert, M. C., Patrick, T. B., Moxley, D. E., et al. (2001). Comparing frequency of content-bearing words in abstracts and texts in articles from four medical journals: An exploratory study. In V. L. Patel, R. Rogers, & R. Haux (Eds.), *MEDINFO 2001: Proceedings of the 10th world congress on medical informatics* (pp. 381–384).

Salton, G., & McGill, M. (1983). *Introduction to modern information retrieval*. New York: McGraw-Hill.

Turney, P. D. (2001). Mining the web for synonyms: PMI-IR versus LSA on TOEFL. In *Proceedings of the 12th European conference on machine learning*.

Van Maanen, L., Borst, J. P., Janssen, C. P., & Van Rijn, H. (2006). Memory structures as user models. In *Proceedings of the 13th annual ACT-R workshop*.

Van Rijn, H., & Anderson, J. R. (2003). Modeling lexical decision as ordinary retrieval. In F. Detje, D. Dörner, & H. Schaub (Eds.), *Proceedings of the fifth international conference on cognitive modeling* (pp. 207–212).