

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/28616>

Please be advised that this information was generated on 2021-10-19 and may be subject to change.

Personalized Search Support for Networked Document Retrieval Using Link Inference

F.C. Berger & P. van Bommel

Computing Science Institute, University of Nijmegen,
Toernooiveld, NL-6525 ED Nijmegen, The Netherlands

Abstract. Constructing a query consisting of a set of terms or descriptors is often an iterative process. To the user, the starting query and the final result could be strongly related. These two queries could even be worthy of a link between them. This paper presents a method for deciding when a link between two descriptors is justified. The decision hinges on the way in which the user has moved from one to the other. In order to allow for users with different levels of experience and different backgrounds, we introduce a number of parameters with which the inference process can be controlled.

1 Introduction

Document retrieval becomes more and more important as the World Wide Web is frequented by searchers from a multitude of backgrounds and with a full spectrum of experience. When a person has to formulate a query in the context of document retrieval, this usually is an iterative process, where to an observer the end result very often only slightly resembles the original query. To the user however these two items are strongly connected. For instance, consider a user who uses an index for constructing a query, and whose search process has started in, say 'hobbies'. After a (perhaps) lengthy sequence of visiting index entries, this searcher ends up in the entry titled 'traditional aspects of Okinawan karate'. When this final entry is tagged by the searcher as relevant, what then can we say about the relation between begin and end of the search process. Although there is no direct relation between the two, to the searcher's point of view there definitely is a correspondence between them. Moreover, they could even be worthy of a link between them.

A fundamental aspect in this is the way in which the searcher moved through the search space of all possible query formulations. When he or she moved in a very erratic way, for instance by moving in circles, then adding a link between beginning and end seems doubtful. Or if contradictory decisions have been made, for instance by first showing a disinterest in the subject of 'martial arts' while later on a definite interest in the subject of 'Okinawan karate' is exhibited. In this case inferring a link between beginning and end seems unwarranted.

This paper presents a formalism for deciding when a search process justifies the introduction of a link. This is far from trivial, since different aspects have to be

considered. For example, the experience level of the searcher plays an important role. Experienced users tend to find too much system support meddlesome, whereas novice users need strong support in order not to become disillusioned with the retrieval system. Our approach is further generalized on the basis of configuration parameters, e.g. several kinds of thresholds. The state of the art in information retrieval is based on hypermedia. There are a number of ways hypermedia based information retrieval can be modeled. We adopt a layered architecture (see e.g. [13]), and specifically a Two-level Hyper-media Architecture (see e.g. [1] [8]).

There are many ways in which a representation of the information need can be constructed. In this paper we adopt a well-known approach called *Query by Navigation* (see e.g. [5] [6] [14] [10] [2]). In Query by Navigation a user is allowed to travel through a hypertext presentation of an index. The aim of this search process is to find a set of document descriptions which is the best description of the information need. The index is constructed through a characterization process in which each document is represented by a hierarchy of broader and narrower phrases in the index. The paper is organized as follows. In Section 2 we introduce the network which is used to navigate. Section 3 deals with the interaction between the searcher and this network. Section 4 treats our view on user profiles, and how these can be used to detect potential links. An important feature of this process of link detection is the feedback given by the user. This feedback is used to fine tune the detection algorithm. The idea of detecting potential links is also the subject of section 5. However in that section we add links based on the way in which the searcher moved through the navigation network. We explain how inconsistent search behaviour can be detected. Finally section 6 briefly highlights the conclusions of this paper. The subject treated in this paper is closely related to the field of data mining and knowledge discovery (see e.g. [15]).

2 Navigation Network

These days a large portion of information retrieval is based on hypermedia. This section treats our view on this current trend. The tool for disclosing a set of documents is a hypermedia-based information retrieval system. The frame of reference in this paper is a two-level hypermedia architecture [7]. This describes how a hypermedia can be formed by creating two levels, the document level or hyperbase, and the index level or hyperindex. Documents can be found in the document level, which is the hypertext representation of a set of documents \mathcal{O} . The content of each document is described by its characterization. The second level, the hyperindex level, is a hypertext representation of the document characterizations. Formally the hyperindex is a tuple $\langle \mathcal{D}, \mathcal{L} \rangle$. The first component is the set of characterization items or descriptors, the second component is the set of links between descriptors.

A main feature of hypertext is the link which connects elements of the hypertext on the basis of some relation between the connected elements. A source for discussion in hypertext research is the nature of this link. One topic is whether links should be unidirectional or bidirectional. We advocate a bidirectional nature of

a link. The main reason for this is that in order to allow query by navigation, it should be possible to reach any descriptor irrespective of the starting-point. So the demand is that if there is a link from d to e , then there should also be a link from e to d . Following Frei & Stieger (see [9]), a link is a tuple $\langle \lambda_T, \lambda_s, \lambda_e, \lambda_c \rangle$ where $\lambda_T \in \mathcal{T}$ is the type of the link, $\lambda_s \in \mathcal{D}$ is the source of the link, $\lambda_e \in \mathcal{D}$ is the destination of the link and $\lambda_c \in \mathbb{N}$ is the activation count of the link. The type of the link determines the relation between source and destination. For instance these two descriptors could be synonyms. For more details, see [4]. For a link λ the source is written as $s(\lambda)$, the destination as $d(\lambda)$, and the activation count as $c(\lambda)$. Each time a link λ is traveled, the activation count λ_c is incremented by 1. As an additional component we could have the date on which it was created, or the date when it was last activated.

3 Query by Navigation

As a means of constructing a query for the retrieval system, Query by Navigation is used. With this formalism a user is allowed to meander through a hypertext representation of an index. Descriptors which are thought to be representative of the Information Need can be marked as such. There is no need for the searcher to navigate the hyperindex with the underlying intention of constructing a query. The sequence of search actions (called a search path) are analyzed and an assumption is made concerning the interests of the searcher. Based on this assumption the search process can be supported by trying to guide the searcher to the index entry which is most likely to be of interest to the searcher. This section discusses the formalism. In order to describe the behaviour of the searcher in the hyperindex we consider four types of actions: (1) mark the focus or one of the options as relevant to the information need (this style of searching is called *berry picking*; see e.g. [3]), which is denoted as $\star\alpha$, (2) shift the focus to one of the options of the current focus, which is denoted as $\rightarrow\alpha$, (3) discard an option as not worthy of further pursuit, which is denoted as $\neg\alpha$, (4) affirming an option as highly interesting for further exploration, which is denoted as $+\alpha$.

Definition 1. Given a search path $p = a_1; \dots; a_k$ we say that descriptor α has been visited when either one of the following events has occurred on p : 1. there is a j such that $a_j = \rightarrow\alpha$; 2. there is a j such that $a_j = \star\alpha$; 3. there is a j such that $a_j = \neg\alpha$; 4. there is a j such that $a_j = +\alpha$

As a simplification of a search path we view a sequence of travelled links and a set of marked descriptors. If a sequence of links $A = \lambda_1; \dots; \lambda_k$ has been travelled, the demand is that we only travel links originating from the descriptor we have just reached by the previous transition.

Lemma 3.1 $\forall_{1 \leq i < k} [s(\lambda_{i+1}) = d(\lambda_i)]$

After the search process has been terminated the searcher could ask the retrieval system which set of documents satisfies the constructed query. The query which will be submitted for evaluation consists of the marked descriptors: $q = \{ \alpha \in \mathcal{D} \mid \text{Mark}(\alpha) \}$

The set of marked descriptors Mark after a sequence of links A has been travelled is a subset of the set of all visited descriptors:

Lemma 3.2 $\text{Mark} \subseteq \bigcup_{i=1}^k d(\lambda_i) \cup \{s(\lambda_1)\}$

The most important characteristics of a search path are its origin, i.e. the descriptor where the search is started, and its destination, i.e. the descriptor where the decision was taken to stop searching because that descriptor offers the best description of the information need. The second most important feature of a search path is its length, i.e. the number of actions which were taken along the path. When we examine these actions, a statement can be made concerning the level of contradiction. This will be high when many actions have been performed which are not consistent with earlier actions. The level of contradiction will be low when no or only a few of such conflicting actions have been performed.

4 Navigation-based User Profile

In this section we address the problem of capturing a searcher's preferences. Given the size of the hyperindex, it is not expected that a user is interested in all descriptors. Most users will travel only a small part of the hyperindex. A good indication of a particular user's preferences is the number of times certain links have been traveled. Uninteresting parts of the hyperindex will be characterized by links which have been traveled only a few times. In contrast, interesting parts will show links which have been activated quite often.

4.1 Activation Count

An important element of a link is the activation count, i.e. the number of times a link λ has been activated. After a searcher has used the hyperindex for some time, certain links will be traveled quite heavily (corresponding to a high activation count), while others are less frequently traveled (corresponding to a low activation count). When viewed as a whole, the set of descriptors being the destination of a link with a high activation count could be said to be a fair representation of the searcher's interest. In order to compare a link's usage to other options, i.e. the *relative* number of times a link emanating from descriptor a has been activated, the normalized activation count is introduced:

Definition 2. The normalized activation count of a link λ originating in descriptor a and going to b is determined by $\bar{c}(\lambda) = \frac{c(\lambda)}{\sum_{\lambda': s(\lambda')=s(\lambda)} c(\lambda')}$

This normalized activation count ranges from 0 up to 1.

A very special class of descriptors is formed by those descriptors being the destination of heavily traveled links, and the source of links which have been traveled very sparsely. This indicates that once a searcher reaches such a descriptor, the motivation for continuing is almost absent. A reasonable cause for this phenomenon is that the descriptor is so relevant that there is no need to continue searching. We shall call such a descriptor a terminating descriptor.

Definition 3. A descriptor d is called a terminating descriptor if there is at least one link λ such that $d(\lambda) = a$ while $c(\lambda)$ is high, and for all links λ' with $s(\lambda') = a$ we have that $c(\lambda')$ is low.

If the normalized activation count of a link λ exceeds a certain significance threshold τ we can say that this link is likely to be traveled.

The question is what happens if a searcher decides to leave the usually traveled paths and starts to travel into a less-traveled part of the hyperindex. In that case the threshold will not be met and these relatively fresh transitions will not be recognized as important. But they could well be very important. So a way has to be implemented to reward a user's investigativeness.

Since an interesting descriptor is marked by the user, the threshold may be lowered if the destination of the link has been marked as relevant to the Information Need. Lowering the threshold is justified because a link may be rewarded for leading to a relevant descriptor. So traveling a link $\langle \lambda_T, a, b, \lambda_c \rangle$ with a low activation count λ_c can be made more important by marking b with $\star b$. This can be seen as a way to assign a novelty value to a transition.

For a given sequence of traveled links A , if each λ_i is reached by following a link with a normalized activation count which exceeds the threshold, we have cause for adding a link from λ_1 to λ_k . In this case the threshold for following a link from λ_i to λ_{i+1} is some function of the distance between λ_1 and λ_i . This function is such that the threshold is raised. The reason for raising the threshold is that we need to demand stronger evidence before labeling an outgoing link as relevant. This leads to the following:

Definition 4. A traveled link λ is called significant if its normalized activation count exceeds the threshold τ , or if its destination is marked or is a terminator:

$$\text{IsSignificant}(\lambda) \Leftrightarrow \bar{c}(\lambda) > \tau \vee d(\lambda) \in \text{Mark} \vee \text{IsTerminator}(d(\lambda))$$

Definition 5. A sequence of links A is called strongly connected if each link λ_i is a significant one: $\text{IsStronglyConnected}(A) \Leftrightarrow \forall_{1 \leq i \leq k} [\text{IsSignificant}(\lambda_i)]$

Suppose a sequence of links is strongly connected. Then any subsequence of this sequence is necessarily also strong connected.

Lemma 4.1 $\text{IsStronglyConnected}(A) \Rightarrow \forall_{1 \leq i < j \leq k} [\text{IsStronglyConnected}(\lambda_i; \dots; \lambda_j)]$

If a searcher has followed a strongly connected sequence of links A_1 and a strongly connected sequence of links A_2 , then the sequence of links $A_1; A_2$ is strongly connected as well, on the condition that A_2 starts in the descriptor which is the destination of A_1 . We now come to the point where, if a strongly connected sequence of links A has been traveled, a link between source and destination of A may be introduced. For such a link we propose a special type, *inferred*.

Definition 6. If a sequence of links A is strongly connected and $s(\lambda_1) \neq d(\lambda_k)$, a link between $s(\lambda_1)$ and $d(\lambda_k)$ is introduced as follows:

$$\mathcal{L} := \mathcal{L} \cup \{ \langle \text{inferred}, s(\lambda_1), d(\lambda_k), 0 \rangle, \langle s(\lambda_k), d(\lambda_1), 0 \rangle \}$$

Because we do not allow reflexive links care needs to be taken not to introduce such links, hence the additional demand that $s(\lambda_1)$ be unequal to $d(\lambda_k)$.

Lemma 4.2 For a strongly connected sequence of links, if no descriptor has been marked, then for each link its normalized activation count must exceed the threshold: $\text{IsStronglyConnected}(A) \wedge \text{Mark} = \emptyset \Rightarrow \forall_{1 \leq i \leq k} [\bar{c}(\lambda_i) \geq \tau]$.

Lemma 4.3 If all visited descriptors have been marked as relevant, then the sequence of links is strongly connected: $\text{Mark} = \bigcup_{i=1}^k d(\lambda_i) \Rightarrow \text{IsStronglyConnected}(A)$.

4.2 Threshold Value

Previously we mentioned that the distance from the start of the sequence of links is a parameter to be used for determining the threshold. Another important parameter is the level of experience of the searcher. A novice searcher is most likely to benefit from a constant threshold, or a threshold which only rises slightly as time goes by. In contrast, an experienced searcher will benefit most from a threshold which either remains at a high level, or increases as time goes by, albeit more quickly than in the case of the novice searcher.

In order to make a statement concerning the informativeness of the number of times a link has been traveled, the *entropy* (see e.g. [11]) of the normalized activation count is used. Given a value x in the interval $[0, 1]$, the entropy $E(x)$ is given by $-x \log x$. A low entropy corresponds with an event of high information value, while a high entropy corresponds with an event of low information value. Because a relatively low traveled and a relatively high traveled link *both* have a low entropy, this measure allows for searchers who decide to travel less-frequently visited areas of the hyperindex. Note that if e_1 and e_2 are two events and e_2 has a higher information value than e_1 , then $E(e_1) > E(e_2)$. Because the amount of needed evidence increases as we move further down the search path, the threshold actually has to be lowered since we now consider entropies. Therefore, $\tau_{i+1} < \tau_i$.

Because we demand more evidence as the distance from the starting point increases, we need to make a statement concerning the maximum threshold value. When we assume that the threshold is based on entropies, we view the case where a link λ is the only one which has been traveled from $s(\lambda)$. In that case, the normalized activation count is 1. This corresponds with entropy $E(\bar{c}(\lambda)) = 0$, i.e. an event which either always occurs or never occurs. Hence, we demand $\lim_{i \rightarrow \infty} \tau = 1$. If the level of experience of a searcher is given by some parameter μ in $[0, 1]$, then $\tau : [0, 1] \times [0, 1] \times \mathbb{N} \rightarrow [0, 1]$, $(\tau_0, \mu, i) \mapsto \tau(\tau_0, \mu, i)$

5 Adding Links Based on Search Behavior

When a searcher uses the hyperindex for the first time, no activation counts are available. Hence we have to add links based on a different source of information. Suppose that the majority of searchers in a hyperindex always travels a path p . Now imagine a searcher μ who decides to travel a deviating path q of length k . The question is when we can introduce a link between $\text{Focus}(q_1)$ and $\text{Focus}(q_k)$ for this user. When q has been constructed in a very erratic way, then clearly the answer is that we may not introduce such a link. On the other hand, when all decisions on q have been made in a consistent way, then adding a link is justified.

5.1 Consistency of Search Paths

In order to make this concept of consistency more clear, suppose we have a searcher who starts searching at the descriptor hobbies. During the search process the user marks the subject martial arts as not relevant. At a later point in the search process the searcher encounters the descriptor titled Okinawan karate. The searcher now decides to mark this descriptor as relevant. Clearly, this is a contradiction, because Okinawan karate is a more specific form of karate, which in turn is a martial art. The introduction of a link (inferred, hobbies, Okinawan karate, 0) (and of course its inverse) is now not justified because of the contradiction. On the other hand, suppose these two decisions occur in reverse order, i.e. first Okinawan karate is marked as not relevant, and some time later martial arts is marked relevant to the information need. In this case a link (inferred, hobbies, Okinawan karate, 0) is justified.

The effect of the conflicting search actions may be disregarded if sufficient time $h \geq 0$ has elapsed between them. This forgetfulness coefficient is a means to incorporate a searcher's level of experience. Novice searchers will benefit from a high forgetfulness (i.e. the contradiction must happen in the few previous actions). Experienced searchers on the other hand should benefit from a low forgetfulness (i.e. the contradiction may occur between search actions which are far apart).

We will only discuss conflicts stemming from marking and rejecting descriptors. In our view, changing focus is merely a way of browsing and hence can not be seen as a search action. In contrast, marking and rejecting are a way to describe the type of subject on which one wants information, and as such are search actions. Affirming a descriptor falls somewhere between these types of interaction.

Definition 7. $\text{IsInconsistent}(a_1; \dots; a_k) \Leftrightarrow \exists_{i,j < i, i-j \leq h} [a_i \perp a_j]$

In this definition $a_i \perp a_j$ is our denotation for conflicting search actions. Conflicts between search actions are commutative, i.e. $a_i \perp a_j \Leftrightarrow a_j \perp a_i$. The decision whether one search action conflicts with another search action can for instance be decided by using situation theory (see e.g. [5]). Another way to decide conflicting search actions is by using some sort of similarity measure between descriptors: $\sigma : \mathcal{D} \times \mathcal{D} \rightarrow [0, 1], (\alpha, \beta) \mapsto \sigma(\alpha, \beta)$ Furthermore we introduce two thresholds: the dissimilarity threshold κ_d and the similarity threshold κ_s , where $0 \leq \kappa_d \leq \kappa_s$. Two descriptors a and b are similar if their similarity measure exceeds the threshold κ_s ; they are called dissimilar if the similarity measure is lower than κ_s . This reflects the Closed World Assumption. Note that $\sigma(\alpha, \alpha) = 1$. We can now decide when two search actions are conflicting. First we treat two search actions dealing with similar search actions. In this case, marking a descriptor, followed by rejecting a similar descriptor can be interpreted as a conflicting sequence: $\sigma(\alpha, \beta) > \kappa_s \Rightarrow \star\alpha \perp \neg\beta$ The conflicting search actions have to occur within the same window h . As an example, suppose we have two synonyms, e.g. holland and the netherlands. Clearly these two descriptors are highly similar. Then, we have $\star \text{holland} \perp \neg \text{the netherlands}$.

Secondly we treat two search actions which deal with dissimilar descriptors. This happens when we first mark a descriptor, followed by marking a dissimilar descriptor: $\sigma(\alpha, \beta) < \kappa_d \Rightarrow \star\alpha \perp \star\beta$ The conflicting search actions have to occur within the same window h . As an example to this situation, suppose we have two antonyms, e.g. monotheism and polytheism. These two descriptors can be seen as highly dissimilar. Then, we have $\star \text{monotheism} \perp \star \text{polytheism}$

Now that we can decide when a sequence of search actions may be called inconsistent, we can discuss when such a sequence may be called consistent.

Definition 8. A sequence of search actions $a_1; \dots; a_k$ is called consistent if according to Definition 7 $a_1; \dots; a_k$ it can not be decided that $a_1; \dots; a_k$ is inconsistent: $\text{IsConsistent}(a_1; \dots; a_k) \Leftrightarrow \neg \text{IsInconsistent}(a_1; \dots; a_k)$

The following lemma is obvious:

Lemma 5.1 If the history $h = 0$, all search paths are consistent.

Definition 8 guarantees consistency within a certain history h . We can also look at *absolute consistency*, i.e. consistency for any history h .

Definition 9. A sequence of search actions $a_1; \dots; a_k$ is called absolute consistent if it can not be proven for any $0 \leq h \leq k$ that $a_1; \dots; a_k$ is inconsistent.

When a sequence of search actions is consistent, any subsequence of these search actions is also consistent.

Lemma 5.2 $\text{IsConsistent}(a_1; \dots; a_k) \Rightarrow \forall_{i,j \geq i} [\text{IsConsistent}(a_i; \dots; a_j)]$

Now we return to the crux of this section: when we have detected a consistent sequence of search actions we wish to add a link between beginning and end of this sequence.

Definition 10. If a sequence of search actions $a_1; \dots; a_k$ is consistent and $\text{Focus}(a_1) \neq \text{Focus}(a_k)$ we may add a link $\langle \lambda_T, \text{Focus}(a_1), \text{Focus}(a_k), 0 \rangle$ to \mathcal{L} .

If we want to include the affirm action into this scheme, we could define the rejection of a previously affirmed descriptor conflicting by using the similarity measure.

5.2 Level of Contradiction

When we have two inconsistent paths, we could be interested in which one is the most inconsistent one. Clearly if a search path is absolute consistent, there is no contradiction. A second component in determining this level of contradiction is how similar the descriptors are which play a part in the conflicting actions. For instance, suppose $\sigma(\alpha, \beta) > \kappa_s$. Then $\star\alpha \perp \neg\beta$, but this conflict is less severe than $\star\alpha \perp \neg\alpha$.

Definition 11. The level of inconsistency ι of a search path is given by

$$\iota = \frac{1 + h - \Delta(h)}{h} \sigma(\text{Subject}(a_i), \text{Subject}(a_j)) \quad \text{where } \Delta(h) = j - i.$$

This level of contradiction can be used in a number of ways. The first of these is in order to make the decision when to add links less severe. In stead of adding only links for consistent search actions, we could choose to add links for search paths up to a certain level of contradiction. The second use is to let the level of contradiction play a part during document retrieval. This is done by multiplying a document's Retrieval Status Value (RSV) with the level of contradiction of the search path on which the RSV's are based.

5.3 Examples

In this section we present an example search path, and use the formalism of section 5.1 and 5.2 to decide on consistency of this search path.

europe; \rightarrow western europe; \neg western europe; \rightarrow the netherlands; \star the netherlands; \rightarrow tourist information
 a_1 a_2 a_3 a_4 a_5 a_6

Fig. 1. Search path

Example 1. Examine the search path of Figure 1, which consists of 6 search actions. Since the Netherlands is a part of Western Europe, we have a conflict between search actions a_3 and a_5 . This renders the search path consistent with $h = 0$ and $h = 1$, but inconsistent with $h = 2$. Thus, the search path is not absolutely consistent.

Example 2. If we choose $h = 3$, the level of inconsistency for the search path of Figure 1 is $\frac{2}{3}\sigma(\text{the netherlands, western europe})$.

Example 3. The expectation is that the number of inconsistent paths increases as the history parameter h is raised. This increase depends on the topology of the hyperindex, viz. the degree in which the descriptors are interconnected, and the number of descriptors which are similar with $\sigma > 0$. Experimental results show that the proportion of inconsistent search paths rises faster for search paths of length 20, than for search paths of length 10.

6 Conclusions and Future Research

In this paper we have presented a mechanism for adding links to a hyperindex. The mechanism caters both for novice searchers by examining their search behavior, and for searchers who have used the hyperindex over a certain period of time. In the case of novice searchers we allow for the effect of time to dampen the occurrence of conflicting search actions. The notion of time also makes its entry in the case of more experienced searchers. There we see that as time elapses

from the starting node of a search path, the amount of evidence needed to label a transition significant increases.

We have introduced several parameters, viz. a significance threshold τ , similarity and dissimilarity thresholds κ_s and κ_d , a user experience level μ and a history threshold h . These parameters may be used to offer a flexible means of support. As a further improvement we may introduce classes of searchers, e.g. a set of classes where a class represents a searcher's background. This class can be used to filter the set of options. Such support is coined *collaborative filtering*. The techniques developed in this paper are also important with respect to a new approach to information retrieval and filtering in the context of the World Wide Web (see e.g. [12]).

References

1. M. Agosti, R. Colotti, and G. Gradenigo. A two-level hypertext retrieval model for legal data. In A. Bookstein, Y. Chiramelia, G. Salton, and V.V. Raghavan, editors, *Proceedings of the 14th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 316–325, Chicago, Illinois, October 1991. ACM Press.
2. H.C. Arents and W.F.L. Bogaerts. Navigation without links and nodes without contents: intensional navigation in a third-order hypermedia system. *Hypermedia*, 5(3):187–204, June 1993.
3. M.J. Bates. The design of browsing and berrypicking techniques for the on-line search interface. *Online Review*, 13(5):407–431, 1989.
4. F.C. Berger and P. van Bommel. Augmenting a Characterization Network with Semantical Information. Technical Report CSI-R9606, Computing Science Institute, University of Nijmegen, Nijmegen, The Netherlands, 1996.
5. F.C. Berger and T.W.C. Huibers. A framework based on situation theory for searching in a thesaurus. *The New Review of Document and Text Management*, 1:253–276, 1995.
6. F.C. Berger and Th.P. van der Weide. A Feedback Mechanism for Query by Navigation. In R. Sacks-Davis and J. Zobel, editors, *Proceedings of the Sixth Australasian Database Conference, ADC'95*, volume 17(2) of *Australian Computer Science Communications*, pages 56–65, Adelaide, Australia, January 1995.
7. P.D. Bruza. *Stratified Information Disclosure: A Synthesis between Information Retrieval and Hypermedia*. PhD thesis, University of Nijmegen, Nijmegen, The Netherlands, 1993.
8. P.D. Bruza and Th.P. van der Weide. Two Level Hypermedia - An Improved Architecture for Hypertext. In A.M. Tjoa and R. Wagner, editors, *Proceedings of the Data Base and Expert System Applications Conference (DEXA 90)*, pages 76–83, Vienna, Austria, 1990. Springer-Verlag.
9. H.P. Frei and D. Stieger. The use of semantic links in hypertext information retrieval. *Information Processing and Management*, 31(1):1–13, 1995.
10. R. Godin, R. Missaoui, and A. April. Experimental comparison of navigation in a galois lattice with conventional information retrieval methods. *International Journal of Man-Machine Studies*, 38(1):747–767, 1993.
11. R.M. Gray. *Entropy and Information Theory*. Springer-Verlag, New York, New York, 1990.
12. E. Hoenkamp, L. Schemaker, P. van Bommel, C.H.A. Koster, and Th.P. van der Weide. Profile - A Proactive Information Filter. Technical Note CSI-N9602, Computing Science Institute, University of Nijmegen, Nijmegen, The Netherlands, 1996.
13. D. Lucarella. A Model for Hypertext-Based Information Retrieval. In *Proceedings of the European Conference on Hypertext - ECHT 90*, pages 81–94, Cambridge, United Kingdom, 1990. Cambridge University Press.
14. J. Nielsen. The art of navigating through Hypertext. *Communications of the ACM*, 33(3):296–310, March 1990.
15. G. Piatesky-Shapiro and editors Frawley, W.J. *Knowledge Discovery in Databases*. AAAI Press, Menlo Park, California, 1991.