Information Coverage in Advisory Brokers

P. van Bommel,* H.A. (Erik) Proper,† Th.P. van der Weide‡
Radboud University, Nijmegen, The Netherlands

Traditionally, information retrieval aims to find information carriers, such as documents, that best match some query or some other (intentional) description of a searcher's information need. In this article, we take the approach that searchers turn to an information retrieval system with the aim of finding several alternatives that completely satisfy their (complex) information need. In other words, searchers expect the retrieval system to help them in covering their information need, rather than merely providing them with a myriad of hopefully relevant information carriers. Ideally, the system should respond by advising one or more packages of information carriers with the requested cumulative effect. This also enables searchers to better trade off between the costs of acquiring and reading/internalizing information carriers versus the expected informational benefits. This article focuses on a theory that aims to clarify the underlying problem area. The theory may consequently be used to enhance information retrieval systems in general and teaching and learning systems in particular, with abilities to better cover a searcher's information need. In the theory presented, we also cater to the fact that searchers may be in different mental moods. The consequence of searchers being in different moods is that information carriers are processed differently. Identifying this influence gives the opportunity to advise users according to their specific moods.

1. INTRODUCTION

The coming of the World Wide Web has opened a new world of human interaction. E-commerce systems can be seen as a commercial exploitation of these new opportunities. The most simple systems offer information structured as hypertext. The interaction with the visitor is very limited. A more general approach is obtained by the introduction of targeted systems and recommender systems. See, for example, Refs. 1–3. Targeted systems are made to be adaptive to visitor behavior, by focusing on correlations between the behavior of visitors. Their goal is to draw the visitor's attention to other objects of interest. Recommender systems go a step further. They assume a set of objects, characterized by their representation. Besides recording visitor's behavior, they use object characterizations and an

*Author to whom all correspondence should be addressed: e-mail: pvb@cs.ru.nl.
†e-mail: erikp@cs.ru.nl.
‡e-mail: tvdw@cs.ru.nl.

INTERNATIONAL JOURNAL OF INTELLIGENT SYSTEMS, VOL. 22, 1155–1188 (2007)
 Published online in Wiley InterScience (www.interscience.wiley.com). • DOI 10.1002/int.20240

expression of the visitor intention. The underlying data model is rather simple, as the main focus of interest for such systems is on matching representation, perceived visitor behavior, and the specification of the visitor interest.

In this article we focus on advanced interaction systems, which we refer to as advisory systems. Rather than a monolithic set of objects, we assume a structured representation of the characterization of information objects in terms of abstract infons. Such advisory systems are particularly suitable for complex information retrieval processes, for example, educational brokers for teaching and learning processes and advisory companies such as intelligent travel agencies. In such complex information retrieval processes, we need mechanisms to derive coherent packages of structured documents, based on underlying knowledge models and user profiles.

The World Wide Web has become the virtual reality of mankind, a world that we shape without many of the imperfections of reality. We can jump to literally every place in no time, live and die many times, change our identity at will, and reach every resource anywhere anytime. In particular this is the promise of information at your fingertips. The growing complexity of information space overwhelms the wired consumer, and the vast increase in information is outpacing the improvement of retrieval tools.

Traditional information retrieval systems aim to satisfy a searcher's information need by matching some explicit formulation of the searcher's information need to the set of available information carriers, such as documents and web pages. The system then returns a set of information carriers that best match the formulation of the searcher's information need. In our opinion, this traditional approach has two serious drawbacks:

Need formulation. Searchers are presumed to have a very clear understanding of their information need, even though it is not likely that they will be able to articulate their precise information needs in terms of a query language. In the case of the Internet, this becomes even more apparent as the collection of available information is endless. Although it has long since been acknowledged (see, e.g., the Cranfield tests[4]) that users have difficulty in expressing this need in a formal language, the fact that searching for information is more of an interactive process of learning, clarification, and discovery is not taken into account. This latter limitation of the information retrieval field is most apparent in the way systems are evaluated. The effectiveness of an information retrieval system is measured in terms of precision and recall for a fixed set of queries on a standardized document collection.

Need satisfaction. Even when the entire set of returned information carriers is indeed relevant to the searcher's information need, a searcher is still required to manually wade through the result sets in search of the right combination of information carriers to cover the information need. The problem with most systems is that a searcher is not provided with advice on an effective combination and order to best read a selection of the information carriers to cover the information need. The word *effectively* here may refer to different aspects, such as time, financial costs, and cognitive load.

In this article, we take the perspective that searchers are not just looking for a set of relevant information carriers, but really expect an advisory system to help them in covering their information need in an effective way. This is what we will refer to as *information coverage*. This research is part of a larger research program on advanced information retrieval techniques. For our other results in this line, see, for example, Refs. 5 and 6.

2. THE INFORMATION COVERAGE PARADIGM

In this section we discuss the information coverage paradigm. We first discuss a general framework for information coverage. Then we focus on contexts for application of the framework.

2.1. Framework

It is our belief that an information retrieval system should really play the role of an information portal, as illustrated by the information coverage paradigm as shown in Figure 1. On the left-hand side there is a searcher who is in need of information. On the right-hand side, we find a collection of available information carriers. This may be a limited collection of information carriers that is available in some library, but could also be an endless set of carriers, such as all information available via the Internet.

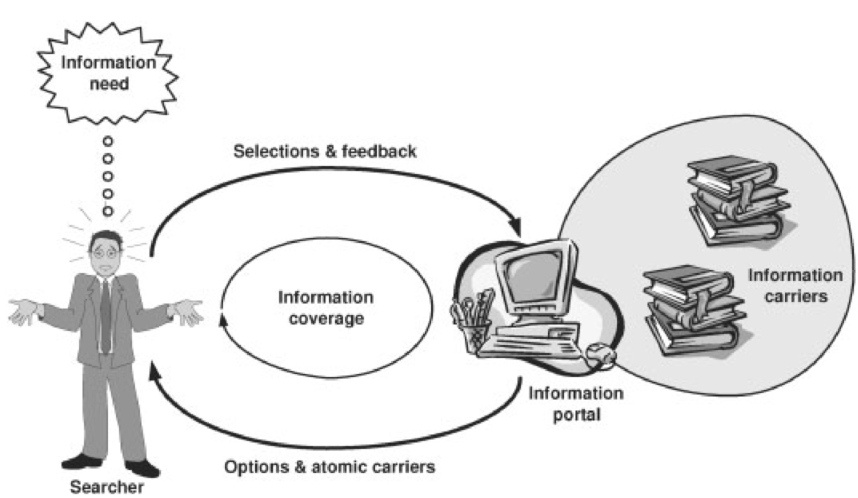

Figure 1. The information coverage paradigm.

In the middle of Figure 1 we find a system, which we will refer to as an *information portal*. The information portal aims to satisfy the searcher's information need. It tries to do so by not only communicating with the searcher in terms of queries and result sets; an information portal should really communicate with searchers on two levels, rather than one:

1. The clarification level focuses on the clarification and discovery of the searcher's actual information needs. At this level of communication, the need formulation issue plays a crucial role.
2. The contents level is concerned with the actual information aimed to fulfill the searcher's information need. At this communication level, the need satisfaction issue is a major challenge.

The communication between the searcher and the information portal should ideally be seen as an interactive dialogue in which the information portal tries to clarify the precise information need while incrementally satisfying this need. Note that special cases of information portals have been proposed, such as recommender systems. Such systems try to make recommendations based on estimations of user preferences. See, for example, Refs. 1–3.

There are many different ways to organize and initiate the communication at the clarification level. For example, in the case of a traditional search engine for the World Wide Web, the onus will be on the searchers, as they are required to express their information need explicitly. In more advanced search engines, the search engine will not only present information carriers that are considered relevant, but may also present possibly related keywords as indications of improved or alternative formulations of the searcher's information need. This enables a more interactive style of communication at the clarification level.

In the case of query by navigation, the onus shifts from the searcher toward the information portal. See, for instance, Refs. 7–9. A query by navigation session would typically start by the information portal offering the searcher a brief list of key topics on which information is available. The searcher may then select a topic from this list, which leads the information portal to react by showing a list of more focused subtopics of the selected topic.

The communication concerning the actual information the searcher is looking for involves the information portal providing the searcher with information carriers and some form of feedback from the searcher to the information portal. Ideally, the feedback from the searcher to the information portal should be as complete as possible with regards to the perceived relevance to the searcher's information needs. The more feedback the information portal receives from the searcher with regards to the relevance of information carriers, the better the information portal will be able to tune the set of selected information carriers to the actual information need. Such a communicative approach is also capable of handling drifting information needs. During the explorative phase of an information coverage process, searchers are bound to learn more about that what they are actually looking for. This is likely to lead to changes (drifts) in their information need.

Note that the two levels of communication cannot be seen separately. When a searcher has finished reading an information carrier from a list of selected information carriers, it is quite likely that the information need of the searcher has changed as a result of the newly gained knowledge. This implies that this searcher will look differently at the relevance of the next information carrier in the list.

The combination of both levels of communication, with the intention of covering a searcher's information need, is the major challenge of information coverage. We aim to develop a theory for this information coverage process, where we take the functionality of an idealized information portal as a starting point. In doing so we will build further on our earlier results.[6] The theory presented there focused on a fundamental grounding of the query by a navigation process. We now aim to extend this theory to formally underpin the information coverage problem.

2.2. Understanding the Searcher

For an information portal, it is most useful to have an understanding of the goal and expectations with which a searcher turns to the system. We consider cognitive and operational levels for defining goals of searchers. The (initial) information need with which a searcher turns to the information portal is the searcher's cognitive goal. In addition to the cognitive goal, a searcher is likely to have some operational goal as well. This goal relates to the tasks that have led the searcher to turn to the system in the hope of gaining new knowledge relevant to the task. For more details about goal-driven learing, see, for example, Ref. 10.

Although, as discussed above, it may not be an easy task for searchers to express their cognitive goal, it will be less hard for them to express their operational goal, especially when this can be done using some predefined terminology/ontology in the context of the searcher's task description, a context that may, for example, be provided by the business process model or a workflow model. Operational goals can have different forms, for instance, task profiles of employees or student goals when attending courses.

The expectations with which a searcher turns to an information portal, in addition to plainly fulfilling their information need, correspond to the cognitive mode of that searcher. A searcher may have different cognitive modes, such as:

1. a breath-first mode where a searcher prefers to be first informed briefly of the different aspects of the subject of their information need
2. a depth-first mode in which searchers prefer to be informed fully about the key issues involved in their information need
3. a repetitive mode where searchers want to be informed about the same issue multiple times, possibly from different perspectives. This mode is particularly useful for searchers who need some time to let new knowledge sink in.
4. a monotonic incremental mode is essentially the opposite of the repetitive mode. In other words, in this mode searchers want to be confronted with as little redundancy as possible with regards to the information they will read.

A single user may actually have different cognitive modes at different points in time. For example, during office hours he or she may operate in a breath-first mode where it concerns information needs pertaining to his or her work tasks, whereas after office hours the same searcher may operate with a depth-first mode when looking for information pertaining to one of his or her hobbies, say.

2.3. Information Coverage Scenarios

Below we briefly discuss two example scenarios of applications of an information portal that aids searchers in the formulation of their information needs and helps them in determining an effective order in which to read selected information carriers. Using these scenarios we would like to argue the socioeconomic relevance of information portals and the research involved. The example scenarios are concerned with personalized information delivery and computer-based training.

2.3.1. Personalized Information Delivery

As our socioeconomic environment progresses from a postindustrial to an information society, the economic climate for content providers, such as publishers, should be favorable. This is, however, not a trivial point. Contemporary Western economies are service based rather than goods based. More and more money seems to be made in providing services than in the selling of goods. In addition, everyone with a PC and Internet connection can be an author and a publisher. Consequently, traditional content providers no longer have a publishing monopoly. This forces the entire content industry into exploring new business models. The emergence of the World Wide Web has raised these questions in the boardroom of every publisher and broadcaster. For more background on these issues, the reader is referred to Refs. 11–14.

The past decennium has already shown some directions in which the ecology of information services may evolve: selling information via the Internet, organizing information in portals, online bookshops that remember individual customers' behavior, multifunctional agents in the form of smartcards, and so on. What these new directions have in common is a combination of new ways of accessing information and innovative business models, used by new entrepreneurial companies that were able to embrace technological advances, thereby reshaping the information ecology. We believe that the above discussed concept of an information portal helping searchers in covering their information needs in an effective way is a potential cornerstone of new business models for content providers. An information portal would provide an excellent means to provide searchers with a personalized access mechanism to the available content. This leads to personalized information delivery. The use of an information portal for personalized information delivery would open up the possibility of providing searchers with the information they would need for the tasks at hand, taking the searchers goals and cognitive modes into proper consideration. The level and quality of personalization would, in addition to the quality and subjects of the actual contents, be the potential competitive advantage a publisher may need.

2.3.2. Computer-Based Training

The importance of knowledge, and in particular the dissemination of knowledge, in modern society does not need any further arguing. In a fast-changing world, organizations will frequently be confronted with the need to disseminate a new body of knowledge within the organization, for example, new knowledge concerning the production of a company's primary product. If available in an electronic format, such bodies of knowledge are likely to be available in a weakly ordered fashion, for example, as computer-based training material, a set of documentation manuals, or a set of web pages.

Different groups within the organization are likely to be interested in different aspects of the body of knowledge as it is available. Decision makers are likely to be interested in different aspects than people from the working floor. Persons from these groups will also have a rather different cognitive identity. Decision makers will be more interested in obtaining an overview of the whole, whereas a work-floor employee is likely to want to be informed about the finest details. A decision maker might get irritated by repeated offering of the same information, whereas another employee would be appreciating this as a valuable feature as supportive for learning.

When disseminating a new body of knowledge in an organization, it would be useful if an information portal would be able to act as a personal mentor to those people who want to acquire portions of the new knowledge. When an information portal would act as a personal mentor, one would expect the information portal to somehow gain an understanding of the learning and reading behavior of the user. When an information portal would exhibit such behavior, it would allow the information portal to better tune the knowledge provided to the individual needs of the users.

2.4. Example Application Context

In this section we describe an example context for the application of information coverage. An employment service office provides the following service. A customer, being interested in a specific job title, is curious to know what educational plan is required to meet the requirements of this job title.

The system is aware of the package of tasks that are associated with this job title. For each task the system knows what knowledge is required to be capable of performing that task. The system starts from the educational level of the customer and determines an educational plan by considering the characterizations of all courses known by the system.

In this example context, we have three levels of description: jobs, tasks, and courses. We suppose the employment service office considers the courses shown in Table I.

Courses are basic information carriers. For each course in Table I, the contents in terms of information items are given. The required foreknowledge is listed as well. As an example, the statement $i_1 \rightarrow i_6$ expresses that item i_6 requires item i_1. We are aware of the fact that the course overview in Table I is rather

Table I. Overview of courses.

Course	Name	Contents	Foreknowledge
DM	Domain modeling	i_1, i_2, i_3	
RE	Requirements engineering	i_4, i_5	
IS	Information systems	i_6, i_7, i_8	$i_1 \to i_6$, $i_2 \to i_6$, $i_4 \to i_7$
TS	Technical systems	i_9, i_{10}, i_{11}	$i_2 \to i_9$, $i_3 \to i_9$, $i_5 \to i_{10}$
SS	System security	i_{12}, i_{13}	
MF	Mathematical foundations	$i_{14}, i_{15}, i_{16}, i_{17}$	
IA	Industrial applications	i_{18}, i_{19}, i_{20}	

simplified and that in practice the required foreknowledge is much more complicated.

Based on course descriptions, a person may be advised to take certain courses. The aim of taking courses is to be capable of performing necessary tasks. We suppose the employment service office considers the tasks shown in Table II.

Tasks may be seen as complex information carriers. For each task in Table II, the required courses are given. Note that, as a consequence of the foreknowledge defined on information items, we may derive that task T_2 requires task T_1. The same holds for task T_3. An overview of jobs is given in Table III.

3. UNDERLYING COST MODELS

What information exactly is has been studied intensively before; see, for example, Refs. 6 and 15. Different authors have provided diverse theories of the nature of information.[16–20] In this article we take a modest approach to information theory, and only assume information to consist of information particles called infons as suggested by Barwise[15,16] and applied to the field of information retrieval by van Rijsbergen and Lalmas,[21] Huibers et al.,[22] and Huibers and Bruza.[23] This broad view on information is in line with the approaches taken in Refs. 16 and 17.

When a collection of alternative information carriers is available to a searcher, some mechanism is needed to determine the relative relevance of these carriers to the searcher's needs. One factor in determining this relevance is the topical relevance of the information carrier with respect to the searcher's information need.

Table II. Overview of tasks.

Task	Name	Courses
T_1	Basics	DM, RE
T_2	Analysis	IS, SS
T_3	Hardware	TS, SS
T_4	Research	MF
T_5	Development	IA

Table III. Overview of jobs.

Job	Name	Required skills
J_1	Information researcher	T_1, T_2, T_4
J_2	Information developer	T_1, T_2, T_5
J_3	Technical researcher	T_1, T_3, T_4
J_4	Technical developer	T_1, T_3, T_5

This is, however, not the only factor involved. Other factors may have to be considered as well. For example,

- the reliability of the information provided by an information carrier
- the amount of money that may have to be paid to obtain the carrier
- the cognitive load the searcher must endure in reading the information carrier
- the amount of time needed to gain access to the information carrier.

To be able to rank the available information carriers while taking such factors into account, a price–performance ratio is introduced. The better the price–performance ratio of an information carrier, the more preferred the carrier is presumed to be.

The actual price–performance ratio associated with a set of information carriers is based on some underlying cost model. As electronic commerce has increasingly gained attention, the need for such models becomes more pressing. For more background about cost models in electronic commerce, see, for example, Ref. 24.

3.1. Nonmonotonicity

Below we will actually see that the price–performance ratio cannot be calculated for one particular information carrier in isolation. Some of the factors involved in computing the price–performance ratio do not behave monotonically with respect to sequences of information carriers. An example of such a factor is the purchase price of an information carrier. When obtaining more information carriers from the same supplier, then the subsequent information carriers may be obtained at a lower price then when they were purchased in isolation. To illustrate this, consider the following example:

For example, suppose chapters A-1 and A-2 are from book A and chapter B-1 is from book B, and A-1 and B-1 are similar in content. Let us also presume book A is priced at $40, book B is priced at $30, and that the searcher does not yet own either book.

If the information portal has to chose from either A-1 or B-1 to fulfill the searcher's information need, the portal is likely to opt for B-1, as this requires the purchase of book B, which is cheaper than book A. However, if A-2 is also needed to cover the searcher's information need, then first purchasing book B might not be such a good idea after all, as it is still required to also purchase book A.

One might argue that in the age of electronic commerce, the above example is antiquated as one in that context one might use micropayments to purchase specific chapters of even smaller pieces of relevant information. However, even in the case of micropayments, the purchase price may still behave nonmonotonically. It is not unlikely that content providers will try and increase the loyalty of their clientele by offering a price reduction based on a customer's volume of consumption. In that case, it may be wiser to purchase chapters A-1 and A-2 using a single content provider, rather than purchasing B-1 and A-2 from different content providers.

Finally, it is not only the purchase price that is likely to behave nonmonotonically. The cognitive load of a searcher is likely to be nonmonotonic as well; when an information portal advises a searcher to first read an introductory article before studying an in-depth report on a certain topic, the cognitive load when reading the in-depth report is likely to be less than when skipping the introductory article.

The cost model considered in this section is defined relative to a certain information need N. This need N is treated as an implicit parameter to the model.

3.2. The Performance of an Information Carrier

The performance of an information carrier is measured in terms of numbers of infons. At present, two factors are identified that contribute positively toward the performance of an information carrier:

- the number of `hits` in terms of the information need covered
- the `curiosity` it raises in terms of infons in the searcher's mind.

Let $\mathcal{PF} = \{\mathtt{hits}, \mathtt{curiosity}\}$ be the set of performance factors used in the cost model. This allows us to define a function $\mathsf{Performance} : \mathcal{PE} \times \mathcal{SE} \times \mathcal{IC} \to \mathbb{R}$ identifying the performance of an information carrier for a given factor and state.

In principle, the number of hits may be determined as $|\mathsf{Hits}(s, N, c)|$. However, the content of information carriers may not always be reliable. In other words, not all of the infons provided by an information carrier may actually be true. Therefore, we presume a function $\mathsf{Reliability} : \mathcal{IC} \to [0, 1]$ to exist, which expresses the reliability of an information carrier in terms of the infons provided by c that are indeed true. Using the notion of reliability, the number of hits provided by an information carrier can be defined as

$$\mathsf{Performance}(\mathtt{hits}, s, c) \triangleq \mathsf{Reliability}(c) \times |\mathsf{Hits}(s, N, c)|$$

The fact that certain infons provided by an information carrier may be false will also be taken into account, however, not as a (negative) performance factor, but rather as a cost factor. Obtaining true infons may come at the cost of having to deal with false infons as well. The performance in terms of the curiosity may simply be defined as

$$\mathsf{Performance}(\texttt{curiosity}, s, c) \triangleq \mathsf{Curiosity}(s, c)$$

We presume the $\mathsf{Performance}$ function to be generalized to sequences of carriers as follows:

$$\mathsf{Performance}(f, s, [c]) \triangleq \mathsf{Performance}(f, s, c)$$

$$\begin{aligned}\mathsf{Performance}(f, s, C ++ D) \triangleq\ & \mathsf{Performance}(f, s, C) \\ & + \mathsf{Performance}(f, s \ltimes C, D)\end{aligned}$$

Depending on the state of the searcher, he or she may value the number of hits provided by an information carrier higher than the curiosity raised or vice versa. In other words, these performance factors need to be weighed in order to combine them into one unified performance value. We therefore presume there to be a weight function:

$$\mathsf{Weight} : \mathcal{PF} \times \mathcal{SE} \rightarrow [0, 1]$$

such that

$$\forall_{s \in \mathcal{SE}} \left[\sum_{f \in \mathcal{PE}} \mathsf{Weight}(f, s) = 1 \right]$$

The overall performance of an information carrier c relative to a searcher S, combining all factors, may then be defined as the weighed average

$$\mathsf{Performance}(s, c) \triangleq \sum_{f \in \mathcal{PE}} \mathsf{Weight}(f, s) \times \mathsf{Performance}(f, s, c)$$

3.3. The Price of an Information Carrier

Information carriers come at a price to the searcher. The price of an information carrier does not only consist of its economic value. There are many different factors involved in the total price. If $\mathcal{PR}$ denotes the set of identified price factors, then the function

$$\mathsf{Price} : \mathcal{PR} \times \mathcal{SE} \times \mathcal{IC} \rightarrow \mathbb{R}$$

is presumed to identify the specific value of a price factor of a given information carrier. The different factors as identified so far of the total cost of an information carrier may be grouped into three phases:

1. First, the searcher must *obtain* the information carrier. The costs of obtaining an information carrier involve three factors:

   - the total (economic) price (`obtain-price`) of obtaining the carrier
   - the amount of time (`obtain-time`) that is required to obtain the carrier
   - the cognitive load (`obtain-load`) on the searcher when obtaining the carrier.

2. The next step is for the searcher to *read* the information carrier. The costs of reading an information carrier involve two factors:

    - the amount of time (`read-time`) that is required to read the carrier
    - the cognitive load (`read-load`) on the searcher when reading the carrier.

3. After the searcher has finished reading an information carrier, there may be additional costs in that the information gleaned from the information carrier may turn out to be *false*. This is captured by the `lies` factor.

This leads to the following set of price dimensions:

$$\mathcal{PR} \triangleq \{\texttt{obtain-price}, \texttt{obtain-time}, \texttt{obtain-load},$$
$$\texttt{read-time}, \texttt{read-load}, \texttt{lies}\}$$

We presume that for `obtain-price`, the Price function also takes into account which information carriers are already purchased by the searcher, as well as any pricing schemes that may be offered by content providers that influence the purchase price.

Using the reliabilty function on information carriers, we are able to define a measure for the cost of being confronted with false information in terms of the number of false infons

$$\mathsf{Price}(\texttt{lies}, s, c) = (1 - \mathsf{Reliability}(c)) \times |\mathsf{Supply}(S, c)|$$

We presume the Price function to be generalized to sequences of carriers as follows:

$$\mathsf{Price}(f, s, [c]) \triangleq \mathsf{Price}(f, s, c)$$

$$\mathsf{Price}(f, s, C ++ D) \triangleq \mathsf{Price}(f, s, C) + \mathsf{Price}(f, s \ltimes C, D)$$

For each of the cost factors we may define the respective price–performance ratios for sequences of informtion carriers as follows:

$$\mathsf{Ratio}(f, s, C) \triangleq \frac{\mathsf{Price}(f, s, C)}{\mathsf{Performance}(s, C)}$$

To be able to combine the three resulting ratios into a single unified price performance, the performance ratios need to be normalized first. This is done by computing the relative deviation of what would be the maximum comfort zone for the searcher with respect to the specific price factor. In other words, what would, on a per-infon basis, be the maximum purchase price at which the searcher would instantaneously approve the purchase of an information carrier? Let the function $\mathsf{MaxPrice} : \mathcal{PR} \times \mathcal{SE} \to \mathbb{R}_{\geq 0}$ denote these maximum values. This allows us to normalize the three price–performance ratios as follows:

$$\mathsf{Ratio}^{\nu}(f,s,C) \triangleq \begin{cases} 0 & \text{if } \mathsf{Ratio}(f,s,C) \leq \mathsf{MaxPrice}(f,s) \\ 1 - \dfrac{\mathsf{MaxPrice}(f,s)}{\mathsf{Ratio}(f,s,C)} & \text{otherwise} \end{cases}$$

mapping them to a domain between 0 and 1. To combine the normalized price/performance ratios into a unified ratio, we combine them using weights:

$$\mathsf{Weight}: \mathcal{PR} \times \mathcal{SE} \rightarrow [0,1]$$

where, again,

$$\forall_{s \in \mathcal{SE}} \left[\sum_{f \in \mathcal{PE}} \mathsf{Weight}(f,s) = 1 \right]$$

This leads to

$$\mathsf{Price}(s,C) = \sum_{f \in \mathcal{PR}} \mathsf{Weight}(f,s) \times \mathsf{Ratio}^{\nu}(f,s,C)$$

4. A THEORY FOR DEMAND AND SUPPLY

Before discussing how information coverage may be achieved in practice, we first need to develop a thorough understanding of the problem area itself. This is done by developing a theory for information coverage, a theory that allows us to define how an idealized information portal should operate in matching demand and supply of information. Once this idealized information portal has been described, we will have a feeling of what to strive for when developing an information portal in practice.

4.1. Information Carriers

Thus far the term *information carrier* has been used without actually providing a definition. In this article, we limit ourselves to information carriers that are available via the Internet. In this context, an information carrier can be described as *any entity from which information may be experienced*. Information carriers may be compound, in which case some mechanism will be available to unfold the carrier into smaller information-carrying objects. Information carriers may be present in several compound information carriers.

Depending on the media used, users can experience an information carrier. For example, users can read the carrier, listen to it, or view it. By stating that a user experiences an information carrier, it is meant that the user reads or views or listens to the information carrier.

4.2. A Model of Searchers

Information carriers can be seen as autonomous objects. However, the amount of information that is actually carried by an information carrier is in the eyes of the beholder. In other words, different searchers will perceive the same information carrier differently. Even more, the same searcher is likely to perceive the same information carriers differently at different moments. For example, when looking for leisure, a searcher will have less appreciation for a theoretical book on mathematics than the next day, when the same searcher is preparing for a mathematical exam. It may also be the case that a searcher who has already experienced an information carrier, when experiencing the same carrier a second time, absorbs additional information that was not absorbed during the first experience. To model this, we take the view that a searcher is in a certain state at the start of experiencing an information carrier and will be in a different state afterward. Let $\mathcal{SE}$ be the set of searcher states. Each state belongs to a unique searcher, determined by the function $\mathsf{Id}: \mathcal{SE} \to \mathcal{ID}$, where $\mathcal{ID}$ is the set of searchers.

To model the subjective perception of information carriers by users, we presume the existence of the following two aspects in the state of a searcher:

- The knowledge a searcher has accrued thus far. This is administrated by the function $\mathsf{Kn}: \mathcal{SE} \to \wp(\mathcal{I})$. The set $\mathcal{I}$ refers to elementary information particles, the *infons* (see Refs. 15 and 16, and for an application to the field of information retrieval see Ref. 6).
- The mood of a searcher. By taking the mood of a searcher into consideration, we are able to model the subjective nature of a searcher experiencing an information carrier. The mood of a searcher in a specific state is obtained by the function $\mathsf{Md}: \mathcal{SE} \to \mathcal{MO}$, where $\mathcal{MO}$ is the set of moods a searcher may have.

If i is the identity of a searcher, then $\mathcal{SE}_i$ is used to denote the set of searcher states of i:

$$\mathcal{SE}_i = \{s \mid \mathsf{Id}(s) = i\}$$

When a searcher experiences an information carrier, then this will lead to a change in both the searcher's knowledge and mood or to a change in state. In this article we restrict ourselves to state changes caused by experiencing information carriers. We do not consider other state changes. For example, forgetting information may be seen as a special change of state. During the interaction between the searcher and the information portal, we assume the state of the searcher to be stable. This is referred to as the *stable searcher assumption*.

We assume searchers to be rational in the sense that their knowledge does not contain gaps. So if a searcher knows some information particle, then this searcher is also aware of the components of this particle. Technically, the knowledge function Kn is presumed to be closed under information inclusion:

[IC1] (Information closure) $\quad \sigma \to \tau \wedge \tau \in \mathsf{Kn}(s) \Rightarrow \sigma \in \mathsf{Kn}(s)$

Here $\sigma \to \tau$ expresses that the information of σ is contained in the information in τ. If a searcher has the intention to acquire some infon τ, then this searcher has

Searchers		
$\mathcal{ID}$	searcher identities	
$\mathcal{SE}$	searcher states	
$\mathcal{MO}$	searcher moods	
Id:	$\mathcal{SE} \to \mathcal{ID}$	identity of searcher
Kn:	$\mathcal{SE} \to \wp(\mathcal{I})$	searcher knowledge in some state
Md:	$\mathcal{SE} \to \mathcal{MO}$	searcher mood in some state
IC1:	***(information closure)***	$g \to f \wedge f \in \mathsf{Kn}(s) \Rightarrow g \in \mathsf{Kn}(s)$
$\mathcal{SE}_i$	$\triangleq \{s \mid \mathsf{Id}(s) = i\}$	

Figure 2. The model of searchers.

information need τ. But in our model, this searcher will also have to learn all preliminary information, that is, all infons σ such that $\sigma \to \tau$. Note that we will not have such a restriction on information carriers. If an information carrier has a knowledge gap, then this gap may be seen as the necessary advance knowledge to be able to experience this carrier. We will come back to this issue in Section 5. Our model of searchers is summarized in Figure 2.

4.3. Information Provision

The information needed by users is provided on information carriers. Formally, information carriers are introduced as the set $\mathcal{IC}$. As stated before, a carrier really only carries data. The carriers are used to transfer information, as data, from one person to another.

For the moment we presume that when a searcher experiences an information carrier, this will be an uninterrupted process where the searcher exclusively concentrates on the specific carrier. In other words, the experience of an information carrier c cannot be interleaved with the experience of another information carrier d. In Section 4.6 we will drop this assumption and discuss compound information carriers and their accumulated effect on the searcher.

When a searcher in state s experiences an information carrier c, then this searcher will end up in a new state denoted as $s \ltimes c$:

$$\ltimes : \mathcal{SE} \times \mathcal{IC} \to \mathcal{SE}$$

Obviously, the new state belongs to the original searcher:

[IC2] (Stable identities) $\mathsf{Id}(s \ltimes c) = \mathsf{Id}(s)$

We assume that searcher knowledge is not lost by experiencing an information carrier. This has been introduced as the stable searcher assumption. The motivation is to safeguard the information portal from a mission impossible.

[IC3] (Nonvolatile searcher memory) $\mathsf{Kn}(s) \subseteq \mathsf{Kn}(s \ltimes c)$

The knowledge of a searcher and the mood this searcher is in are determining for the observing potential of this searcher.

[IC4] (Base for observing) Let s_1 and s_2 be states of the same searcher; then $\mathsf{Kn}(s_1) = \mathsf{Kn}(s_2) \wedge \mathsf{Md}(s_1) = \mathsf{Md}(s_2) \Rightarrow \mathsf{Kn}(s_1 \ltimes c) = \mathsf{Kn}(s_2 \ltimes c)$.

By this axiom, the functions Kn and Md materialize the cognitive identity of a searcher. An immediate consequence is that differences in information absorption must be explained by a different mood of the searcher.

LEMMA 1. *Let s_1 and s_2 be states of some the same searcher, then*

$$\mathsf{Kn}(s_1) = \mathsf{Kn}(s_2) \wedge \mathsf{Kn}(s_1 \ltimes c) \neq \mathsf{Kn}(s_2 \ltimes c) \Rightarrow \mathsf{Md}(s_1) \neq \mathsf{Md}(s_2)$$

As we remarked, a searcher may absorb additional information when experiencing an information carrier a second time. Therefore, we will *not* assume $\mathsf{Kn}((s \ltimes c) \ltimes c) = \mathsf{Kn}(s \ltimes c)$. The $\ltimes$ operator is left-associative, and thus we may omit parentheses and write $s \ltimes c_1 \ltimes c_2$ rather than $(s \ltimes c_1) \ltimes c_2$.

Example 1. Assume {interested, informed, bored, alerted} $\subseteq \mathcal{MO}$; then repeatedly offering the same document may lead to

1. $\mathsf{Md}(s) = \text{interested}$
2. $\mathsf{Md}(s \ltimes c) = \text{informed}$
3. $\mathsf{Md}(s \ltimes c \ltimes c) = \text{bored}$.

The rationale for this is the following rule of experience:

$$\mathsf{Md}(s) = \text{bored} \Rightarrow \forall_c[\mathsf{Kn}(s) = \mathsf{Kn}(s \ltimes c)]$$

which states that bored people do not learn. In the absence of external state change triggers for state change (the stable searcher assumption), there can only be an informative way to escape from being bored:

$$\mathsf{Md}(s \ltimes c \ltimes c \ltimes \text{agenda}) = \text{alerted}$$

4.3.1. Exploring the Search Space

The experience operator induces an ordering relation on searcher states. State s_1 may result in state s_2 (denoted as $s_1 \twoheadrightarrow^* s_2$) by experiencing information carriers subsequently. This relation is the reflexive transitive closure of experiencing a single information carrier:

[IC5] (Learning trajectories) $s_1 \twoheadrightarrow^* s_2 \Leftrightarrow s_1 = s_2 \vee \exists_c [s_1 \ltimes c \twoheadrightarrow^* s_2]$

The condition $\phi \twoheadrightarrow^* \psi$, where ϕ and ψ are predicates over states, is introduced as

$$\phi \twoheadrightarrow^* \psi \triangleq \forall_{s_1}[\phi(s_1) \Rightarrow \exists_{s_2}[\psi(s_2) \wedge s_1 \twoheadrightarrow^* s_2]]$$

Example 2. Let ϕ be a predicate; then we introduce

$$i \in \phi \triangleq \exists_s[\phi(s) \wedge \mathsf{Id}(s) = i]$$

We call ϕ a group predicate if

$$\forall_s[\mathsf{Id}(s) \in \phi \Rightarrow \phi(s)]$$

The group thus covers states $\{s | \phi(s)\}$ and has members $\{i | i \in \phi\}$. We call ϕ a subtype of ψ if

$$\phi \sqsubseteq \psi \triangleq \forall_s[\phi(s) \Rightarrow \psi(s)] \wedge \psi \text{ is a group predicate}$$

As an example,

$$\text{starting student} \sqsubseteq \text{student}$$

A subtype will have actual group knowledge, but may also be supposed to satisfy some knowledge threshold.

Example 3. Assuming courses are the information carriers provided, a teaching program must supply sufficient information to go through subsequent stages:

1. starting student $\twoheadrightarrow^*$ 1st year student
2. 1st year student $\twoheadrightarrow^*$ 2nd year student
3. 2st year student $\twoheadrightarrow^*$ bachelor.

These conditions between state predicates must hold for all individual students.

We assume each searcher has an initial state upon entering the information portal. Let $b(i)$ be the initial state of searcher i.

[IC6] (Entry point) $b(\mathsf{Id}(s)) \twoheadrightarrow^* s$

New states can only be reached by experiencing an information carrier. The finiteness of a searcher experience history is expressed by the *State Induction Scheme*.

4.3.2. Personal and Collective Carrier Semantics

The supply of an information carrier to the searcher knowledge is expressed as the incremental information provision (see Figure 3) of an information carrier in a given searcher state:

$$\mathsf{Supply}(s,c) \triangleq \mathsf{Kn}(s \ltimes c) - \mathsf{Kn}(s)$$

The supply of an information carrier may vary from state to state. Each state may extract some special kind of infons. Yet it will be useful to realize the overall information content of an information carrier for a given searcher. For this purpose, the information semantics of an information carrier, relative to a searcher i, is defined as the potential information it may provide to this searcher:

$$\mathsf{InfoSem}(c,i) \triangleq \bigcup_{s \in \mathcal{SE}_i} \mathsf{Supply}(s,c)$$

The above definition is taken from the searcher's point of view. From the information carrier's point of view, the informational semantics will be introduced as the result of the indexing process, that is, the characterization of the information carrier. In this section, we restrict ourselves to the viewpoint of the searcher. The general informational semantics, the *infomantics*, of an information carrier may be defined as the whole of information it may provide to any searcher:

$$\mathsf{InfoSem}(c) \triangleq \bigcup_{i \in \mathcal{ID}} \mathsf{InfoSem}(c,i)$$

LEMMA 2. $\mathsf{InfoSem}(c) = \cup_s \mathsf{InfoSem}(s,c)$.

Personal information semantics		
$\ltimes$:	$\mathcal{SE} \times \mathcal{IC} \to \mathcal{SE}$	experience searcher of infons
IC2:	***(stable identities)***	$\ltimes(s_1,c) = s_2 \Rightarrow \mathsf{Id}(s_1) = \mathsf{Id}(s_2)$
IC3:	***(Nonvolatile searcher memory)***	$\mathsf{Kn}(s) \subseteq \mathsf{Kn}(s \ltimes c)$
IC4:	***(Base for observing)***	$\mathsf{Kn}(s_1) = \mathsf{Kn}(s_2) \wedge \mathsf{Md}(s_1) = \mathsf{Md}(s_2)$ $\Rightarrow \mathsf{Kn}(s_1 \ltimes c) = \mathsf{Kn}(s_2 \ltimes c)$
IC5:	***(Learning trajectories)***	$s_1 \twoheadrightarrow^* s_2 \Leftrightarrow s_1 = s_2 \vee \exists_c [s_1 \ltimes c \twoheadrightarrow^* s_2]$
Supply(s,c)	$\triangleq$ $\mathsf{Kn}(\ltimes(s,c)) - \mathsf{Kn}(s)$	
$\phi \twoheadrightarrow^* \psi$	$\triangleq$ $\forall_{s_1} [\phi(s_1) \Rightarrow \exists_{s_2} [\psi(s_2) \wedge s_1 \twoheadrightarrow^* s_2]]$	
InfoSem(c,i)	$\triangleq$ $\bigcup_{s \in \mathcal{SE}_i} \mathsf{Supply}(s,c)$	
InfoSem(c)	$\triangleq$ $\bigcup_i \mathsf{InfoSem}(c,i)$	

Figure 3. A model for information provision.

Proof.

$$\begin{aligned}\sigma \in \mathsf{InfoSem}(c) &\Leftrightarrow \exists_i[\sigma \in \mathsf{Supply}(c,i)]\\ &\Leftrightarrow \exists_i\exists_s[\mathsf{Id}\ (s) = i \wedge \sigma \in \mathsf{Supply}(c,i)]\\ &\Leftrightarrow \exists_s[\sigma \in \mathsf{Supply}(s,c)]\end{aligned}$$ ∎

4.3.3. *Comparing Information Absorption*

When experiencing the same information carrier in different searcher states, the supplied information will differ. This allows us to compare different states of the same searcher based on his or her capacity for absorbing information from that information carrier. We will call a state s_1 subabsorbing to state s_2 if the information supplied by each carrier c is also present after experiencing c from state s_2:

$$\mathsf{SubAbsorber}(s_1,s_2) \triangleq \forall_c[\mathsf{Supply}(s_1,c) \subseteq \mathsf{Kn}(s_2 \ltimes c)]$$

The situation is depicted in Figure 4.

Example 4. Let s_1 be a state where the searcher is bored; then each other state of this searcher will be absorbing. The reason is then being bored ($\mathsf{Md}(s)$ = bored implies to no information carrier has a learning effect: $\forall_c[\mathsf{Kn}(s) = \mathsf{Kn}(s \ltimes c)]$. Thus for each carrier c we have $\mathsf{Supply}(s,c) = \emptyset$ and, consequently, $\mathsf{SubAbsorber}(s,s')$ for each other state s'.

This absorption relation is a partial order on searcher states.

LEMMA 3. *The relation* SubAbsorber *is both reflexive and transitive.*

The absorbing potential is further described by the following lemma.

LEMMA 4. *Let* $\mathsf{SubAbsorber}(s_1,s_2)$*; then* $\mathsf{Kn}(s_1) \subseteq \mathsf{Kn}(s_2) \Rightarrow \mathsf{Kn}(s_1 \ltimes c) \subseteq \mathsf{Kn}(s_1 \ltimes c)$.

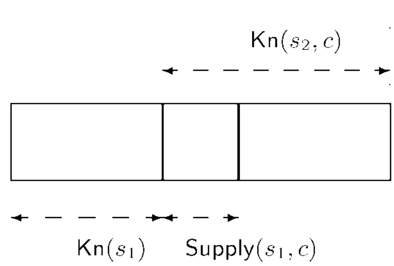

Figure 4. $\mathsf{SubAbsorber}(s_1,s_2)$.

Proof. Suppose $\mathsf{SubAbsorber}(s_1, s_2)$, and let $\mathsf{Kn}(s_1) \subseteq \mathsf{Kn}(s_2)$; then

$$\begin{aligned}\mathsf{Kn}(s_1 \ltimes c) &= \mathsf{Kn}(s_1) \cup \mathsf{Supply}(s_1, c)\\ &\subseteq \mathsf{Kn}(s_1) \cup \mathsf{Kn}(s_2 \ltimes c)\\ &\subseteq \mathsf{Kn}(s_2) \cup \mathsf{Kn}(s_2 \ltimes c)\\ &= \mathsf{Kn}(s_2 \ltimes c)\end{aligned}$$ ■

We do not assume SubAbsorber to be an antisymmetric relation. We will further explore this situation. States s_1 and s_2 are considered equally absorbing ($\mathsf{EqAbsorber}(s_1, s_2)$) if they are mutually subabsorbing:

$$\mathsf{EqAbsorber}(s_1, s_2) \triangleq \mathsf{SubAbsorber}(s_1, s_2) \wedge \mathsf{SubAbsorber}(s_2, s_1)$$

The motivation for this definition is the generalization of Lemma 4.

LEMMA 5. *Let* $\mathsf{EqAbsorber}(s_1, s_2)$*; then* $\mathsf{Kn}(s_1) = \mathsf{Kn}(s_2) \Rightarrow \mathsf{Kn}(s_1 \ltimes c) = \mathsf{Kn}(s_1 \ltimes c)$.

So, in equally absorbing states, the searcher mood has no effect on the assimilating potential of knowledge. The situation is depicted in Figure 5.

LEMMA 6.

$$\mathsf{InfoSem}(c, i) = A \cup B \cup C$$

where

- $A = \mathsf{InfoSem}(c, i) - \mathsf{Supply}(s_1, c)$ *is the knowledge in information carrier c that searcher i is not familiar with in state* s_1.
- $C = \mathsf{InfoSem}(c, i) - \mathsf{Supply}(s_2, c)$ *is the analogon for state* s_2.
- $B = \mathsf{Supply}(s_1, c) \cap \mathsf{Supply}(s_2, c)$.

Proof. First we note that both $\mathsf{Supply}(s_1, c) \subseteq \mathsf{InfoSem}(c, i)$ and $\mathsf{Supply}(s_2, c) \subseteq \mathsf{InfoSem}(c, i)$ due to the definition of InfoSem. Next let σ be an infon from

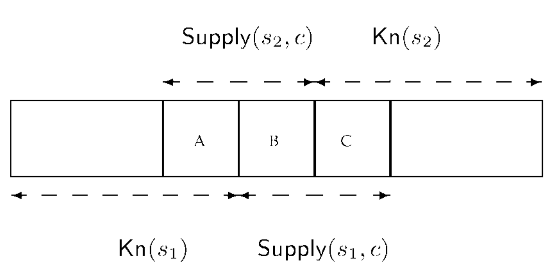

Figure 5. An impassive searcher.

$\mathsf{InfoSem}(c,i)$ that is not available in either A or C. As $\sigma \notin A$, we conclude $c \in \mathsf{Supply}(s_1,c)$. Analogously we conclude $\sigma \in \mathsf{Supply}(s_2,c)$. This enables us to conclude $\sigma \in B$. ■

4.3.4. The Impassive Absorber

A typical searcher that may be distinguished is referred to as the *impassive absorber*: a searcher who will always obtain the same information from an information carrier, independent of the mood the searcher is in:

$$\mathsf{ConsistentAbsorber}(i) \triangleq \forall_{s_1,s_2 \in \mathcal{SE}_i}[\mathsf{EqAbsorber}(s_1,s_2)]$$

This is, for example, the case for an automated searcher, as such a searcher cannot change moods. Searchers who have a constant mood are impassive.

LEMMA 7. $\forall_{s_1,s_2 \in \mathcal{SE}_i}[\mathsf{Md}(s_1) = \mathsf{Md}(s_2)] \Rightarrow \mathsf{ConsistentAbsorber}(i)$.

Proof. Axiom IC4 provides the cognitive identity of a searcher by the functions Kn and Md. Let s_1 and s_2 be states of the same searcher; then from this axiom we conclude that $\mathsf{Kn}(s_1) = \mathsf{Kn}(s_2) \wedge \mathsf{Md}(s_1) = \mathsf{Md}(s_2) \Rightarrow \mathsf{Kn}(s_1 \ltimes c) \neq \mathsf{Kn}(s_2 \ltimes c)$. As $\mathsf{Md}(s_1) = \mathsf{Md}(s_2)$, we have $\mathsf{Kn}(s_1) = \mathsf{Kn}(s_2) \Rightarrow \mathsf{Kn}(s_1 \ltimes c) \neq \mathsf{Kn}(s_2 \ltimes c)$. Consequently, $\mathsf{ConsistentAbsorber}(i)$. ■

By this axiom, the functions Kn and Md provide the cognitive identity of a searcher. An immediate consequence is that differences in information absorption must be explained by a different mood of the searcher. See Figure 6.

LEMMA 8. *Let s_1 and s_2 be states of the same searcher; then* $\mathsf{Kn}(s_1) = \mathsf{Kn}(s_2) \wedge \mathsf{Kn}(s_1 \ltimes c) \neq \mathsf{Kn}(s_2 \ltimes c) \Rightarrow \mathsf{Md}(s_1) \neq \mathsf{Md}(s_2)$.

An impassive user is an effective information consumer.

LEMMA 9. *If i is an impassive absorber with state s, then*

$$\mathsf{InfoSem}(c,i) \subseteq \mathsf{Kn}(s \ltimes c)$$

$\mathsf{SubAbsorber}(s_1,s_2)$	$\triangleq \forall_c\,[\mathsf{Supply}(s_1,c) \subseteq \mathsf{Kn}(s_2 \ltimes c)]$
$\mathsf{EqAbsorber}(s_1,s_2)$	$\triangleq \mathsf{SubAbsorber}(s_1,s_2) \wedge \mathsf{SubAbsorber}(s_2,s_1)$
$\mathsf{ConsistentAbsorber}(i)$	$\triangleq \forall_{s_1,s_2 \in \mathcal{SE}_i}\,[\mathsf{EqAbsorber}(s_1,s_2)]$

Figure 6. A model to compare information absorption.

Proof. Let i be an impassive absorber with state s; then for each other state s' of this searcher we have $\mathsf{Supply}(s', c) \subseteq \mathsf{Kn}(s \ltimes c)$, which leads to

$$\begin{aligned}\mathsf{InfoSem}(c, i) &= \bigcup_{s' \in \mathcal{SE}_i} \mathsf{Supply}(s', c) \\ &\subseteq \bigcup_{s' \in \mathcal{SE}_i} \mathsf{Kn}(s \ltimes c) \\ &= \mathsf{Kn}(s \ltimes c)\end{aligned}$$

■

There is another way to put this.

LEMMA 10. *If i is an impassive absorber, then*

$$\mathsf{Id}(s) = i \Rightarrow \mathsf{Kn}(s) \cup \mathsf{InfoSem}(c, i) = \mathsf{Kn}(s \ltimes c)$$

As a consequence, an impassive absorber is a maximal absorber. The information that is obtained by an impassive absorber does not depend on the mood of this kind of searcher. An impassive absorber will fully commit all information that can be obtained from an information carrier to memory. In other words, we have the following lemma.

LEMMA 11. *Let s be a state of an impassive absorber i; then* $\mathsf{Kn}(s \ltimes c \ltimes c) = \mathsf{Kn}(s \ltimes c)$.

4.4. Cognitive Features of Searchers

The *incremental model* for information retrieval is based on an increment function that measures the residual relevance $I(S, x)$ of a document x after a set S of documents already has been presented to this searcher. The set S is also referred to as the miniprofile of the searcher. In this article we focus on the Supply function, which measures the incremental provision of an information carrier. We will discuss the cognitive searcher features presented in Ref. 5 in terms of the model presented in this article.

The cognitive feature *Repetition* describes the effect of experiencing an information carrier for a second time. A searcher has this feature if a second experience is completely redundant:

$$\text{IM1} \quad \textit{Repetition:} \quad s_1 \ltimes c \twoheadrightarrow^* s_2 \Rightarrow \mathsf{Supply}(s_2, c) = \emptyset$$

The cognitive feature *Growth* expresses a monotonic effect of searcher knowledge on information recording capability of the searcher:

$$\text{IM2} \quad \textit{Growth:} \quad \mathsf{Kn}(s_1) \subseteq \mathsf{Kn}(s_2) \Rightarrow \mathsf{Supply}(s_1, c) \supseteq \mathsf{Supply}(s_2, c)$$

The following cognitive features describe how information carriers influence each other for each searcher. We will say that information carrier c is about carrier d, denoted as $c \preccurlyeq d$, when experiencing d makes c superfluous for searcher i:

$$c \preccurlyeq_i d \triangleq \mathsf{Kn}(b(i) \ltimes d \ltimes c) = \mathsf{Kn}(b(i) \ltimes d)$$

A searcher possesses the cognitive feature *Effectiveness* when the relation $\preccurlyeq$ is transitive:

$$\text{IM3} \quad \textit{Effectiveness:} \quad c \preccurlyeq_i d \wedge d \preccurlyeq_i e \Rightarrow c \preccurlyeq_i e$$

A weaker version of this cognitive feature is named *Effective growth*:

$$\text{IM3a} \quad \textit{Effective growth:} \quad x \preccurlyeq y \Rightarrow \forall_S [I(S, x) \leq I(S, y)]$$

Information carrier c is independent of carrier d when the experience of d does not affect the experience of c. This is denoted as $c \rfloor d$, in that

$$c \rfloor_i d \triangleq \mathsf{Supply}(b(i) \ltimes c, d) = \mathsf{Supply}(b(i), d)$$

The cognitive feature *Independence* states that this property is carried over to all states (see Table IV):

$$\text{IM4} \quad \textit{Independence:} \quad c \rfloor_i d \Rightarrow \forall_{s \in \mathcal{SE}_i} [\mathsf{Supply}(s \ltimes c, d) = \mathsf{Supply}(s, d)]$$

The last cognitive feature describes the relation between aboutness and independence:

$$\text{IM5} \quad \textit{Exclusion:} \quad c \rfloor_i d \wedge e \preccurlyeq_i d \Rightarrow c \rfloor_i d$$

Using these cognitive features, searcher classes may be defined. For example:

The globe trotter. The first searcher class we consider is the *globe trotter*, examining a particular field of interest in order to find out sufficient details. In terms of the search process, a globe trotter is seen as a searcher trying to cover some topic of interest without really being interested in completeness. Experiencing new sensations is the incentive of this searcher.

The student. Next we consider the cognitive setting of the *student*. A student is a searcher who is trying to get acquainted with some topic. The topic is not stable; reading a document might draw the student's attention to a new area of interest. Reading an information object a second time may be profitable, especially when documents read in between have contributed knowledge that enables the student to learn more in a second reading pass.

The collector. A rather different searcher class is the *collector*. A collector is a searcher wishing to collect information objects with respect to some topic.

Table IV. Cognitive identities characterized.

Searcher type	Repetition	Growth	Effectiveness	Independence	Exclusion
	×	—	—	—	—
Globe trotter	—	×	×	—	—
Student	—	—	—	×	×
Collector	×	×	×	×	×

It is not profitable to have an object more than once. The collector tries to make the collection complete.

A collective searcher does not benefit from experiencing the same document twice, even if this searcher has experienced other information carriers in between:

$$\mathsf{Collecting}(i) \triangleq \forall_{s_1, s_2 \in \mathcal{SE}_i}[s_1 \ltimes c \twoheadrightarrow^* s_2 \Rightarrow \mathsf{Supply}(s_2, c) = \emptyset]$$

LEMMA 12. *An impassive searcher is a collective searcher.*

Proof. Let i be an impassive searcher with states s_1 and s_2 such that $s_1 \ltimes c \twoheadrightarrow^* s_2$. From this latter property, we conclude $\mathsf{Kn}(s_1 \ltimes c) \subseteq \mathsf{Kn}(s_2)$. As i is impassive, we have $\mathsf{Supply}(s_2, c) \subseteq \mathsf{Kn}(s_1 \ltimes c)$. Consequently, $\mathsf{Supply}(s_2, c) \subseteq \mathsf{Kn}(s_2)$. This is only possible if $\mathsf{Supply}(s_2, c) = \emptyset$. ■

4.5. The Demand of a Searcher

A searcher may have several needs. For example, in our ongoing example, a searcher may have a need for computer science, English literature, and art history. The set of all information needs is denoted as $\mathcal{IN}$. In each searcher state each need is present to some extent. In the context of this article, we will restrict ourselves to the need for information.

The information need a searcher may have corresponds to a need for infons. What infons are needed depends on the current mood of the searcher and what the searcher already knows. This subjective meaning of information need is modeled by the function

$$\mathsf{Demand} : \mathcal{SE} \times \mathcal{IN} \to \wp(\mathcal{I})$$

The total need $\mathsf{Need}(s)$ of a searcher in some state s thus is obtained by

$$\mathsf{Need}(s) = \sum_N \mathsf{Demand}(s, N)$$

Differences in the meaning of a searcher's information need are either caused by differences in a searcher's mood or in her or his level of knowledge:

[IC7] (Effect of mood) Let s_1 and s_2 be two states of the same searcher; then,

$$\mathsf{Demand}(s_1, N) \neq \mathsf{Demand}(s_2, N)$$
$$\Rightarrow \mathsf{Md}(s_1) \neq \mathsf{Md}(s_2) \vee \mathsf{Kn}(s_1) \neq \mathsf{Kn}(s_2)$$

Searchers do not cheat in their actual demand for infons; in other words, knowledge that is already known is not needed:

[IC8] (No cheating) $\mathsf{Kn}(s) \cap \mathsf{Demand}(s, N) = \emptyset$

A consequence is that it does not make sense to offer an impassive absorber the same information twice. Another impact of this axiom is that experiencing an information carrier has an influence on the subjective meaning of the information need.

LEMMA 13. $\mathsf{Demand}(s \ltimes c, N) \cap \mathsf{Supply}(s, c) = \emptyset$.

Proof. Using the definition of Supply we have

$$\begin{aligned}
&\mathsf{Demand}(s \ltimes c, N) \cap \mathsf{Supply}(s, c)\\
&\quad = \mathsf{Demand}(s \ltimes c, N) \cap (\mathsf{Kn}(s \ltimes c) - \mathsf{Kn}(s))\\
&\quad \subseteq \mathsf{Demand}(s \ltimes c, N) \cap \mathsf{Kn}(s \ltimes c)\\
&\quad = \emptyset
\end{aligned}$$

■

4.5.1. *Classifying Information Exchange*

The relevance of an information carrier c with regards to a need N for a searcher $\mathsf{Id}(s)$ can be expressed as follows:

$$\mathsf{Overhead}(s, N, c) \triangleq \mathsf{Supply}(s, c) - \mathsf{Demand}(s, N)$$

$$\mathsf{Shortage}(s, N, c) \triangleq \mathsf{Demand}(s, N) - \mathsf{Supply}(s, c)$$

$$\mathsf{Hits}(s, N, c) \triangleq \mathsf{Supply}(s, c) \cap \mathsf{Demand}(s, N)$$

This situation is shown in Figure 7. Hits represent the benefits of an information carrier. Overhead information will be negatively appreciated. Shortage corresponds to unfulfilled searchers' needs. However, information shortage may not be part of the subjective meaning of the information need anymore, as a result of experiencing the information carrier.

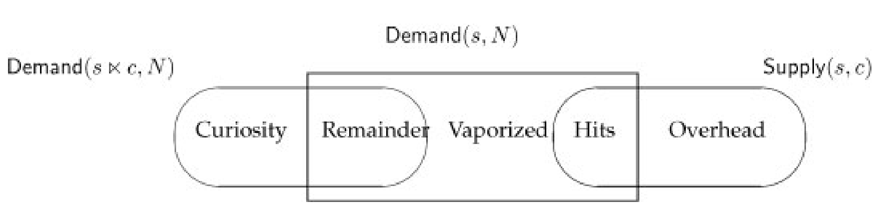

Figure 7. Demand/supply classes.

The demand for information, given a certain information need N, will also be influenced by experiencing information carriers:

$$\mathsf{Satisfaction}(s, N, c) \triangleq \mathsf{Demand}(s, N) - \mathsf{Demand}(s \ltimes c, N)$$

$$\mathsf{Remainder}(s, N, c) \triangleq \mathsf{Demand}(s, N) \cap \mathsf{Demand}(s \ltimes c, N)$$

$$\mathsf{Curiosity}(s, N, c) \triangleq \mathsf{Demand}(s \ltimes c, N) - \mathsf{Demand}(s, N)$$

Satisfaction corresponds to that part of the information need that has been satisfied by experiencing an information carrier. In our model, the presentation of an information carrier is a reason for searcher satisfaction. This is expressed by the following lemma.

LEMMA 14. $\mathsf{Hits}(s, N, c) \subseteq \mathsf{Satisfaction}(s, N, c)$.

Proof. Let $f \in \mathsf{Hits}(s, N, c)$; then (1) $f \in \mathsf{Supply}(s, c)$, leading to $f \notin \mathsf{Demand}(s \ltimes c, N)$ (Lemma 13), and (2) $f \in \mathsf{Demand}(s, N)$. Consequently $f \in \mathsf{Satisfaction}(s, N, c)$. ■

The residual information need is part of the missing information.

LEMMA 15. $\mathsf{Remainder}(s, N, c) \subseteq \mathsf{Shortage}(s, N, c)$.

Proof. Let $f \in \mathsf{Remainder}(s, N, c)$; then (1) $f \in \mathsf{Demand}(s, N)$ and (2) $f \in \mathsf{Demand}(s \ltimes c, N)$, leading to $f \in \mathsf{Supply}(s, c)$ (Lemma 13). Consequently $f \in \mathsf{Shortage}(s, N, c)$. ■

Searcher satisfaction cannot be a result of overhead information. This is an immediate consequence of the definition of these quantities.

LEMMA 16. $\mathsf{Satisfaction}(s, N, c) \cap \mathsf{Overhead}(s, N, c) = \emptyset$.

Axiom IC8 asserts that the knowledge supplied by an information carrier will no longer be part of the information need of the searcher. We take an even stronger position and presume that searcher satisfaction is only caused by the information carrier presented.

[IC9] (Causal satisfaction) $\mathsf{Satisfaction}(s, N, c) \subseteq \mathsf{Supply}(s, c)$

The consequence of this axiom is that searchers are impassive, in the sense that they will not lose interest in information shortage.

LEMMA 17. $\mathsf{Shortage}(s, N, c) \subseteq \mathsf{Demand}(s \ltimes c, N)$.

Proof. Let $f \in \mathsf{Shortage}(s,N,c)$; then (1) $f \in \mathsf{Demand}(s,N)$ and (2) $f \notin \mathsf{Supply}(s,c)$, and thus $f \notin \mathsf{Satisfaction}(s,N,c)$ (Axiom IC9), which can be rewritten as $f \notin \mathsf{Demand}(s,N) \vee f \in \mathsf{Demand}(s \ltimes c,N)$. Combining (1) and (2) leads to $f \in \mathsf{Demand}(s,N) \wedge f \in \mathsf{Demand}(s \ltimes c,N)$. ∎

If an information carrier is exhaustive for an information need, then this information carrier contains all requested information.

LEMMA 18. $\mathsf{Demand}(s \ltimes c,N) = \emptyset \Rightarrow \mathsf{Demand}(s,N) \subseteq \mathsf{Supply}(s,c)$.

Proof. Suppose $\mathsf{Demand}(s \ltimes c,N) = \emptyset$; then Axiom IC9 yields $\mathsf{Demand}(s,N) \subseteq \mathsf{Supply}(s,c)$. ∎

4.5.2. Demand Drift

Inspecting an information carrier may lead to an interest in new information. This is referred to as *demand drift*. In our model (see Figure 8), demand drift is captured in the function $\mathsf{Curiosity}(s,N,c)$. Typically, during a search session, the information portal will assume a nondrifting searcher. However, in other applications, for example a teaching environment, the intention of the information portal might be to bring about a specific demand drift, in order to provoke the student's interest in relevant materials (of course, the information portal might also try to bring the student in a beneficial mood).

First we investigate the absence of demand drift. An immediate consequence of the definition of Curiosity is the following lemma.

Information		
$\mathcal{IN}$	Information Needs	
Demand:	$\mathcal{SE} \times \rightarrow \wp(\mathcal{I})$	information need of searcher
IC8: ***(no cheating)***		$\mathsf{Kn}(s) \cap \mathsf{Demand}(s,N) = \varnothing$
IC7: ***(Effect of mood)***		$\mathsf{Demand}(s_1,N) \neq \mathsf{Demand}(s_2,N) \Rightarrow \mathsf{Md}(s_1) \neq \mathsf{Md}(s_2) \vee \mathsf{Kn}(s_1) \neq \mathsf{Kn}(s_2)$
IC9: ***(Causal satisfaction)***		$\mathsf{Satisfaction}(s,N,c) \subseteq \mathsf{Supply}(s,c)$
Overhead(s,N,c)	$\triangleq$	$\mathsf{Supply}(s,c) - \mathsf{Demand}(s,N)$
Shortage(s,N,c)	$\triangleq$	$\mathsf{Demand}(s,N) - \mathsf{Supply}(s,c)$
Hits(s,N,c)	$\triangleq$	$\mathsf{Supply}(s,c) \cap \mathsf{Demand}(s,N)$
Satisfaction(s,N,c)	$\triangleq$	$\mathsf{Demand}(s,N) - \mathsf{Demand}(\ltimes(s,c),N)$
Remainder(s,N,c)	$\triangleq$	$\mathsf{Demand}(s,N) \cap \mathsf{Demand}(\ltimes(s,c),N)$
Curiosity(s,N,c)	$\triangleq$	$\mathsf{Demand}(\ltimes(s,c),N) - \mathsf{Demand}(s,N)$

Figure 8. The model of demand.

LEMMA 19. *If* $\mathsf{Curiosity}(s,N,c)=\emptyset$*, then* $\mathsf{Demand}(s \ltimes c,N) \subseteq \mathsf{Demand}(s,N)$. *Furthermore, in that case* $\mathsf{Remainder}(s,N,c) = \mathsf{Demand}(s \ltimes c,N)$*, and thus* $\mathsf{Demand}(s \ltimes c,N) \subseteq \mathsf{Shortage}(s,N,c)$.

For such search actions, the residual information need amounts to the shortage of the presented document.

LEMMA 20. *If* $\mathsf{Curiosity}(s,N,c)=\emptyset$*, then* $\mathsf{Demand}(s \ltimes c,N) \subseteq \mathsf{Shortage}(s,N,c)$.

Proof. If $f \in \mathsf{Demand}(s \ltimes c,N)$, then from the premise and Lemma 19, it follows that $f \in \mathsf{Demand}(s,N)$. It furthermore follows from Lemma 13 that $f \notin \mathsf{Supply}(s,c)$. We may therefore conclude that $f \in \mathsf{Demand}(s,N) - \mathsf{Supply}(s,C)$. From the definition of $\mathsf{Shortage}$, then it immediately follows that $f \in \mathsf{Shortage}(s,N,c)$. ■

Lemmas 17 and 20 may now be combined into the following lemmas.

LEMMA 21. *If* $\mathsf{Curiosity}(s,N,c)=\emptyset$*, then* $\mathsf{Demand}(s \ltimes c,N) = \mathsf{Shortage}(s,N,c)$.

LEMMA 22. *If* $\mathsf{Curiosity}(s,N,c)=\emptyset$*, then* $\mathsf{Remainder}(s,N,c) = \mathsf{Shortage}(s,N,c)$.

Proof. From the definition of $\mathsf{Remainder}$ it follows that

$$\mathsf{Remainder}(s,N,c) = \mathsf{Demand}(s,N) \cap \mathsf{Demand}(s \ltimes c,N)$$

With Lemma 21 we have

$$\mathsf{Remainder}(s,N,c) = \mathsf{Demand}(s,N) \cap \mathsf{Shortage}(s,N,c)$$

As $\mathsf{Demand}(s,N) \subseteq \mathsf{Shortage}(s,N,c)$, we can conclude that

$$\mathsf{Remainder}(s,N,c) = \mathsf{Shortage}(s,N,c)$$ ■

4.6. Compound Information Carriers

So far, information carriers have been considered to be elementary. In this section we focus on compound information carriers (see Figure 9). There will be two possibilities to construct new information carriers. The first operator, called *sequencing*, concatenates information carriers. Let c_1 and c_2 be information carri-

$+ :$	$\mathcal{IC} \times \mathcal{IC} \to \mathcal{IC}$	Sequential presentation
IC10:	***(Sequential decomposition)***	$s \ltimes (c_1 + c_2) = s \ltimes c_1 \ltimes c_2$
$\parallel :$	$\mathcal{IC} \times \mathcal{IC} \to \mathcal{IC}$	Parallel presentation

Figure 9. Compound information carriers.

ers; then their concatenation is denoted as $c_1 + c_2$. Concatenated information carriers are experienced by the searcher in the order in which they are concatenated.

[IC10] (Sequential decomposition) $s \ltimes (c_1 + c_2) = s \ltimes c_1 \ltimes c_2$

At this point we conclude that compound information carriers can describe the relation $\twoheadrightarrow^*$ easily:

LEMMA 23. $s_1 \twoheadrightarrow^* s_2 \Leftrightarrow \exists_c [s_1 \ltimes c = s_2]$.

The second constructor for information carriers lets the searcher experience the involved information carriers at the same time. This is denoted as $c_1 \| c_2$. The interpretation is that information carrier c_2 supports the effect of c_1.

[IC] (Parallel decomposition) $\mathsf{Kn}(s \ltimes (c_1 \| c_2)) \subseteq \mathsf{InfoSem}(c_1)$

Information carrier c_2 is said to *support* c_1, with respect to a searcher state s, iff

$$\mathsf{Kn}(s \ltimes c_1) \subseteq \mathsf{Kn}(s \ltimes (c_1 \| c_2))$$

It is said to *disturb* c_1 with respect to s iff

$$\mathsf{Kn}(s \ltimes (c_1 \| c_2)) \subseteq \mathsf{Kn}(s \ltimes c_1)$$

5. A REASONING MODEL FOR INFORMATION PORTALS

Before a learning trajectory is started, a learner is in a certain state $s_1 \in \mathcal{SE}$. This state can be thought of as an encapsulation of the learner's current mood, knowledge, and identity. His knowledge at this point is $\mathsf{Kn}(s_1)$. Note that this is a subset of the total amount of information that exists in the world: $\mathsf{Kn}(s_1) \subseteq \mathcal{I}$.

By learning, we want to increase our knowledge. The only way to learn is to study information carriers. Recall that Axiom IC4 states that learning an infon by definition increases our knowledge. Learning, and thus extending our knowledge, is not enough in the context of our example. We want to achieve a certain task, so we should learn the *right* things. This means that we want to move from our current state s_1 to a specified state s_n in which the learning task is completed. Even more so, we want to do this as efficiently as possible by optimizing the steps from s_1 to s_n. Note that $\mathsf{Kn}(s_1) \subseteq \mathsf{Kn}(s_n) \subseteq \mathcal{I}$.

5.1. Toward Learning Packages

The central problem we consider here is the following: given initial state s and information need N, determine a cost-optimal compound information carrier from the set of carriers that satisfy this need:

$$\{c \in \mathcal{IC} | \mathsf{Demand}(s \ltimes c, N) = \emptyset\}$$

The rationale is that information carriers will contribute to searcher knowledge. However, a searcher may not be capable of grasping information available in an information carrier, as the searcher lacks the required base knowledge. This problem is caused by the Information Closure rule expressed in Axiom IC1. Let $\mathsf{Pre}(i)$ be the required foreknowledge for infon i:

$$\mathsf{Pre}(i) = \{i' \mid i' \rightarrow i \wedge i' \neq i\}$$

Infon i thus can only be supplied by information carrier c in searcher state s if the required foreknowledge is available in this state:

$$i \in \mathsf{Supply}(s, c) \Rightarrow \mathsf{Pre}(i) \subseteq \mathsf{Kn}(s) \cup \mathsf{Ch}(s)$$

The foreknowledge for information carrier c thus consists of all infons that are not mentioned in this carrier, but are contained in an infon that is part of carrier c:

$$\mathsf{Pre}(c) = \{i' \mid i' \notin \mathsf{Ch}(c) \wedge \exists_{i \in \mathsf{Ch}(c)}[i' \rightarrow i]\}$$

An information carrier is appropriate for a searcher in state s if

$$\mathsf{IsFit}(s, c) \equiv \mathsf{Pre}(c) \subseteq \mathsf{Kn}(s)$$

5.2. Learning Packages

A series of information carriers is referred to as a learning package. Consider learning package $c_1, \ldots, c_k$. The foreknowledge $\mathsf{Pre}(c_i)$ of package c_i should be available after the student has processed carriers $c_1, \ldots, c_{i-1}$. Assume the student starts in state s; then, when entering on carrier c_i, the knowledge of the student amounts to

$$\mathsf{Kn}(s \ltimes c_1 \ltimes \cdots \ltimes c_{i-1})$$

The missing knowledge at this point for a successful complete processing (i.e., *optimal processing*) of carrier c equals

$$\mathsf{Pre}(c_i) - \mathsf{Kn}(s \ltimes c_1 \ltimes \cdots \ltimes c_{i-1})$$

Thus the total foreknowledge that is required for the learning package $c_1, \ldots, c_k$ for optimal processing is expressed as

$$\mathsf{Pre}_{\mathsf{subj}}(c_1, \ldots, c_k) = \bigcup_{i=1}^{k} (\mathsf{Pre}(c_i) - \mathsf{Kn}(s \ltimes c_1 \ltimes \cdots \ltimes c_{i-1}))$$

This is referred to as the *subjective approach*. It requires a thorough insight into the knowledge of a student and the effect of experiencing information carriers.

The *objective approach* is based on the objective estimation of the information contents of an information carrier. This approach is prescriptive, in the sense that it requires the student to obtain the normative information contents for a successful participation in a learning package. Upon processing information carrier c_i, the student should have learned (and not forgotten!)

$$\sum_{j=1}^{i-1} \mathsf{Ch}(c_j)$$

So the student comes short of

$$\mathsf{Pre}(c_i) - \sum_{j=1}^{i-1} \mathsf{Ch}(c_j)$$

and thus the required knowledge for optimal processing of the learning package may be identified as

$$\mathsf{Pre}_{\mathsf{obj}}(c_1, \ldots, c_k) = \bigcup_{i=1}^{k} \left(\mathsf{Pre}(c_i) - \sum_{j=1}^{i-1} \mathsf{Ch}(c_j) \right)$$

After processing the learning package, starting from state s, the student will be in state $s \ltimes c_1 \ltimes \cdots \ltimes c_k$. In subjective terms, this determines the resulting level of the student:

$$\mathsf{Post}_{\mathsf{subj}}(s) = \mathsf{Kn}(s \ltimes c_1 \ltimes \cdots \ltimes c_k)$$

In objective terms, the student is assumed to have reached level

$$\mathsf{Post}_{\mathsf{obj}}(s) = \mathsf{Kn}(s) \cup \bigcup_{i=1}^{k} \mathsf{Ch}(c_i)$$

5.3. Providing Advice

Suppose a student is interested in a course that is characterized by required knowledge level PRE, training the student onto knowledge level POST. Furthermore, the course requires the student to be in mood m. The initial state s of the student then should be such that $\mathsf{PRE} \subseteq \mathsf{Kn}(s)$ and $\mathsf{Md}(s) = m$.

A semisubjective approach to the effect of an information carrier is an averaging schema for what a student may learn form this carrier. The extreme values are determined by the results of the best students and the information that is obtained by the average student:

$$\bigcap_{s:\mathsf{Md}(s)=m} \mathsf{Supply}(s, c) \subseteq \mathsf{InfoSem}(c, m) \subseteq \bigcup_{s:\mathsf{Md}(s)=m} \mathsf{Supply}(s, c)$$

Let $\mathsf{Cost}(s, c)$ be the cost for optimal processing information carrier c in state s. Then the total cost of a learning package is expressed as

$$\mathsf{Cost}_s(c_1, \ldots, c_k) = \sum_{i=1}^{k} \mathsf{Cost}(s \ltimes c_1 \ltimes \cdots \ltimes c_{i-1}, c_i)$$

5.4. Examples of Learning Packages

A learner called Lou is faced with the task of learning something. (S)he has to learn about several topics including discrete mathematics, economics, and econometrics in order to get a certificate in economics. In this example, Lou has to take one final exam that covers all the topics that (s)he learned about before.

Lou recognizes that some of the topics overlap. For example, "functions" are likely to be covered in both mathematics and econometrics, whereas the cost models are likely to be covered in both economics and econometrics. The amount of time available for study implies that the student must use his/her time efficiently, which in turn implies that studying the same subject many times should be avoided whenever possible. This is where the recommender system comes in handy: it will advise Lou which material should be read and in what order.

Lou logs into the system which "automagically" knows everything about Lou. At this point the system knows his/her mood, background knowledge, and which well-defined task must be completed. The system is now challenged with the task to give advice on which material should be read and in what order. In order to do this the system must know the following:

- the beginning state Lou is in, which encompasses both mood and knowledge (a set of infon's)
- the desired end state after the learning trajectory (also a set of infon's)
- the infons that are contained in the study material as well as the required knowledge before any infon can be studied and understood
- what effect learning an infon has on Lou's mood.

There are several possibilities for the recommendation that the system comes up with. Assuming that there is exactly one book to be read for each topic, two examples of recommendations are:

1. Read the mathematics book first, then study economics, and finally use what is learned to understand the economics book more quickly.
2. Start by reading the econometrics book. To be able to understand a section about calculation price elasticity, some deeper insights in both economics and mathematics must be gained first. So, the "normal track" is interrupted and Lou has to read sections of the other two books at this point.

By reading the recommended literature in the specified way, Lou's knowledge will grow incrementally. Assuming that the system made a good outline, Lou won't have to read the same material over and over. Also, the system makes sure that Lou has enough prior knowledge before each chunk of information (infon) is learned. In the end, Lou should know enough to receive the certificate.

6. CONCLUSIONS AND FURTHER RESEARCH

In this article we considered a general model for information coverage. Advisory systems should not only give relevant documents to searchers, but should help searchers in covering their information need effectively. To be able to build such advisory systems, we need a flexible notion of *cost*. We described such a notion, but at this point we do need to examine the applicability of our interpretation of cost in more detail.

In our theory of demand and supply, we proposed a number of basic rules (axioms) for information coverage (IC axioms). The theory was further augmented with basic properties in the form of lemmas, which are derivable from IC axioms. Note that information carriers may support and disturb each other. This has been defined in terms of the knowledge operator Kn in Section 4. We need to examine similar but other forms of influence between information carriers.

Further research is directed toward the specialization of our theory to specific situations and application domains, and to employ advanced techniques (such as Bayesian belief networks and advanced cost models) to equip advisory systems with better reasoning abilities.

Future advisory systems should help nonexperts in complex tasks more intelligently. We do believe that concepts such as learning packages are a main challenge in this respect. Although our first experiences are positive, a more elaborate practical validation is necessary.

References

1. Terveen L, Hill W, Amento B, McDonald D, Creter J. PHOAKS: A system for sharing recommendations. Communication of the ACM 1997;40:59–62.
2. Sarwar BM, Karypis G, Koning H, Riedl J. Item-based collaborative filtering recommendation algorithms. In: 10th Int Proc World Wide Web Conf, Hong Kong, May 2001. pp 285–295.
3. Herlocker JL, Konstan JA, Riedl J. Explaining collaborative filtering recommendations. In: Proc Conf on Computer Supported Cooperative Work; 2000. pp 241–250.
4. Cleverdon CW. The significance of the Cranfield Tests on index languages. In Bookstein A, Chiarmarella Y, Salton GE, Raghavan VV, editors. Proc 14th Annual ACM Conf of Research and Development in Information Retrieval (SI-GIR'1991), Chicago, IL, October 1991. New York: ACM Press; 1991. pp 3–12.
5. van der Weide ThP, Huibers TWC, van Bommel P. The incremental searcher satisfaction model for information retrieval. Comput J 1998;41:311–318.
6. Proper HA, Bruza PD. What is information discovery about? J Am Soc Inform Sci 1999;50:737–750.
7. Bruza PD, van der Weide ThP. Stratified hypermedia structures for information disclosure. Comput J 1992;35:208–220.
8. Bruza PD, van der Weide ThP. The modelling and retrieval of documents using index expressions. ACM SIGIR Forum 1991;25(2):6–25.
9. Bruza PD, van der Weide ThP. Two level hypermedia—An improved architecture for hypertext. In: Tjoa AM, Wagner RR, editors. Proc Data Base and Expert System Applications Conference (DEXA 90), Vienna, Austria, Berlin, Germany. Berlin: Springer; 1990. pp 76–83.
10. Ram A, Leake D. A framework for goal-driven learning. In: des Jardins M, Ram A, editors. Proc 1994 AAAI Spring Symp on Goal–Driven Learning; 1994. pp 1–11.
11. Tapscott D. Digital economy—Promise and peril in the age of networked intelligence. New York: McGraw-Hill; 1996.
12. Tapscott D, Caston A. Paradigm shift—The new promise of information technology. New York: McGraw-Hill; 1993.

13. Davenport TH. Process innovation: Reengineering work through information technology. Boston, MA: Harvard Business School Press; 1993.
14. Davenport TH, Short JE. The new industrial engineering: Information technology and business process redesign. Sloan Management Review 1990(Summer):11–27.
15. Devlin K. Infons and types in an information-based logic. In: Cooper R, Mukai K, Perry J, editors. Situation theory and its applications—Volume 1. Stanford, CA: Center for the Study of Language and Information; 1990. pp 79–96.
16. Barwise J. The situation in logic. Stanford, CA: Center for the Study of Language and Information; 1989.
17. Landman F. Towards a theory of information. Dordrecht: Foris; 1986.
18. Losee RM. The science of information. New York: Academic Press; 1990.
19. Losee RM. A discipline independent definition of information. J Am Soc Inform Sci 1997;48:254–269.
20. Shannon CE. A mathematical theory of communication. Bell Syst Techn J 1948;22: 379–423,623–656.
21. van Rijsbergen CJ, Lalmas M. Information calculus for information retrieval. J Am Soc Inform Sci 1996;47:385–398.
22. Huibers TWC, Lalmas M, van Rijsbergen CJ. Information retrieval and situation theory. SIGIR Forum 1996;30:11–25.
23. Huibers TWC, Bruza PD. Situations: A general framework for studying information retrieval. In: Leon R, editor. Information retrieval: New systems and current research. Proc 16th Research Colloquium of the British Computer Society Information Retrieval Specialists Group. Drymen, UK: Taylor Graham; 1996. pp 3–25.
24. Wondergem BCM, van Bommel P, Huibers TWC, van der Weide ThP. Opportunities for electronic commerce in information discovery. In Griffel F, Tu T, Lamersdorf W, editors. Proc Int IFIP/GI Working Conf on Trends in Distributed Systems for Electronic Commerce, TrEC 98, June 1998, pp 126–136.