Thesis

A User-Driven Adaptive Interaction Strategy To Support Exploratory Searching

561

Colofon

Document: Master Thesis
Version: 1.0
Date: 09/08/2007

Student 1
Name: Jeroen Bakker
Studentnr: 0421898
Thesisnr: 561
E-mail: J.Bakker@student.ru.nl

Institution: Radboud Universiteit Nijmegen
Programme: Informatica
Department: Information Retrieval & Information Systems (IRIS)
Supervisor: Prof. dr. ir. Theo van der Weide
2nd Teacher: Prof. dr. Mario van Vliet
Address: Toernooiveld 1, 6525 ED Nijmegen

Abstract

Most Information Retrieval models offer only limited support to searchers who possess only a vague and inexpressible understanding of their actual information need. A search process that begins with this inexplicit understanding is referred to as exploratory searching. This research aims to improve the current state of research surrounding exploratory searching, by providing the searcher with additional support throughout the search process.

Prior to presenting our vision to address this dilemma, some fundamental concepts within Information Retrieval are revisited and expanded upon, based on insights derived from established research within this field.

The presented model for Information Retrieval aims to facilitate exploratory searching by supporting the searcher through adaptive interaction, precipitating the searcher's increasingly accurate understanding of his information need.

To illustrate the defined concepts and models in this thesis, they are finally discussed in the light of an example domain.

Keywords: Awareness Development Paths, Consolidated Intermediation Model, Exploratory Search, Human Computation, Information Behaviour, Information Need Awareness, Information Need Levels, Information Retrieval Intermediation, Intermediation Strategies, Search Behaviour, System Awareness, System Awareness Levels

Preface

I would like to take this opportunity to thank my research partner, Daniël Rutten, for his input in our master thesis research project that spanned the past few months. Working with Daniël was once again a blast, resulting in quite some interesting insights!

Furthermore, the both of us would like to thank Theo van der Weide and Mario van Vliet of the Information Retrieval & Information Systems (IRIS) group for their supervision and support. Additionally, there are several people that have helped us in many different ways, ranging from discussing the research to providing moral support. We would like to thank everyone that helped us in any way.

Table of Contents

1. Introduction

"The mere existence of information does not ensure access to it" - Thomas Mann

"A User-Driven Adaptive Interaction Strategy To Support Exploratory Searching" is the result of a master thesis project executed by Daniël Rutten and Jeroen Bakker, two computer science graduate students at the Radboud University of Nijmegen, under the supervision of prof. Theo van der Weide and prof. Mario van Vliet. This research is keen on improving the state of current research on information retrieval and (in particular) laying a foundation for effective support of exploratory searching. The structure of this document is as follows:

- Chapter 2 covers the *Problem Statement* and includes the research questions.
- An analysis of *Information Behaviour* is provided in chapter 3, which defines the terminology and related concepts used in this research, in order to establish a common vocabulary and eliminate ambiguity.
- Chapter 4 deals with *Search Behaviour*, exhibited by the searcher during his interaction with Information Systems.
- In chapter 5, *Practical Intermediation Models* are considered as a source of inspiration and best-practices of everyday intermediation scenario's, in which a searcher is guided in his quest for information.
- Chapter 6 offers a *Consolidated Intermediation Model* based upon the aforementioned intermediation models, and is designed to assist and guide a searcher in fulfilling his need for information during the exploratory search process.
- Chapter 7 puts the *Theory in Practice*, by projecting the findings on an example domain, thereby providing some more insight in their use.
- The information needs from the example domain are put to the test in chapter 8: *Fulfilling the Information Needs*.
- The *Conclusion* of our research and the contribution it offers to the field Information Retrieval is reflected upon in chapter 9.
- Additional *Recommendations,* to further expand upon this research, are offered in chapter 10.

This research builds on and attempts to incorporate multi-disciplinary research done in the fields of Information Retrieval (IR), Information Seeking and Human-Computer Interaction (HCI). As such, it encompasses various perspectives on IR, including the searcher's cognitive-behavioural aspects as well as interface design and user-adaptation.

2. Problem Statement

"The best way to escape a problem, is to solve it."
- Alan Saporta

This chapter discusses the research questions. Some background information is also provided to help establish the context of the research.

2.1. Background

In everyday life, people are faced with problems and situations that need to be overcome by making a concious decision. Typically, the person's initial approach would be to resolve the situation by using his personal knowledge to oversee the problem. However, when this knowledge fails to provide the person with a reasonable approach to the situation, he will have to obtain more knowledge, to get from his current state of knowledge to the desired state of knowledge. The difference between these two states is known as the *knowledge gap*. Bridging this gap is accomplished by having the person assume the role of a searcher, searching for and assimilating information. In today's society, however, there is an abundance of information available to anyone who is interested in it. This virtually unlimited availability of large quantities of information describing any number of topics in varying depths, introduces a serious dilemma: how does the searcher manage to locate the information most relevant to him?

The goal of the searcher's searching behaviour, is to acquire the appropriate knowledge for bridging his knowledge gap. Typically, there are several aspects to consider when 'building' this bridge. For instance, the searcher can ask someone else for information, or try to find this information in one of many information objects (books, papers, articles, etc.), made available by an information repository, known as an Information System.

However, the bigger the knowledge gap, the more effort it takes to bridge this gap. Furthermore, when the problem space is unclear and ill-defined, the exact specifications for the bridge are vague. The searcher's ability to express his need for information depends on the searcher's perception of his knowledge gap and his intrinsic degree of uncertainty. As such, searchers do not always succeed in accurately formulating their information need. Taylor has done some pioneering work, identifying several stages of a searcher's *information need awareness*, referring to them as *information need levels* [TAYLOR1968]; the ability to express one's need depends on the information need awareness they are subjected to.

2.2. Research Questions

Many search-engines and -techniques currently exist; most of them, however, offer only a single form of interaction and do not adapt their strategies to the searcher's inherent information need awareness. Although most existing Information Systems are quite effective when searchers are in a high state of information need awareness, they offer little to no support to searchers with a low information need awareness. As such, the guidance offered to the searcher in the IR process is not necessarily optimal and is subject to improvement. Therefore, this study will focus on *exploratory searching* (see *paragraph 4.1: Exploratory Searching*), in which the searcher is initially unable to formulate a precise query. This suggests that some form of guidance and support during the search process would act as a catalyst for further refining the searcher's information need.

The main research question that can be derived from the aforementioned problem is as follows:

> ***Main Research Question**: "How can a searcher best be assisted in refining his information need throughout the process of exploratory searching, aimed towards guiding the searcher to the most relevant resources available?"*

In order to ensure effective communication throughout this paper, a common foundation of IR (including vocabulary and semantics) is required to remove any ambiguity of the concepts in our research. This leads to the first research question:

> ***Research Question 1**: "Which models are used to represent various aspects of Information Retrieval and how do the concepts defined therein relate to one another?"*

To assess how best to assist people during their search process, it is important to understand the nature of exploratory searching and how people search for information in the first place. The following research question can therefore be formulated:

> ***Research Question 2**: "Which search strategies do people employ to individually fulfil their information need, with respect to their intrinsic information need awareness?"*

To this end we will analyse and compare several key models of the search process in the field of cognitive science and information retrieval, in order to establish the current state of affairs of the information searching process.

In addition to the (commonly unassisted) search strategies derived from research question 2, searchers can also be assisted in their search process by an *intermediary*. To define the strategies that an intermediary can employ, inspiration is acquired from everyday situations in which people are assisted via human-to-human intermediation. This results in research question 3:

> ***Research Question 3**: "Which practical intermediation models are employed in everyday life, to support searchers in their quest to fulfil their information needs?"*

By using the insights derived from the practical intermediation models, devising a new (consolidated) intermediation model for Information Retrieval may be possible. Therefore, the following research question can be formulated:

> ***Research Question 4**: "Can the insights derived from the practical intermediation models be translated to the field of digital information retrieval and, if so, can a (consolidated) intermediation model be created?"*

Aside from the main research question, an individual research goal is postulated. It relates to applying the findings to an example domain, to see how the model could hold in practice.

> ***Individual Research Goal**: "Project the findings on an example domain and investigate the added value."*

Towards this end, the following sub-questions are posed:

> ***Research Question 5**: "In the example domain, how is the problem space experienced and which types of information need exist?"*

To answer this question, the chosen example domain is introduced, with the problem space context being evaluated in this light.

> ***Research Question 6**: "How do the Awareness Development Paths relate to searching behaviour in the example domain?"*

For this question, the relation of information need awareness and system awareness is presented in terms of the example domain.

> ***Research Question 7**: "In the example domain, how do Information Systems support searchers satisfying their information needs?"*

This question relates to the information and applicable information paths offered by information systems in the domain. The information and corresponding internal characteristics will be related to the information needs from the previous question, in order to establish the applicability of the IS for the information needs.

> ***Research Question 8**: "In the example domain, does the Consolidated Intermediation Model provide additional support over the established Information Systems?"*

After relating the Awareness Development Paths to the example domain, this question conveys the benefits of the Consolidated Intermediation Model over the 'traditional' approach?

3. Information Behaviour

"Information is not knowledge." - Albert Einstein

In order to establish an unambiguous understanding of IR, some concepts (and inherent relationships) will first be defined, to allow for a unified context throughout this document.

3.1. Information Representation

Information Retrieval pertains to the process of a searcher requesting for information, followed by the searcher's consumption of that information. Once consumed (i.e. read) and interpreted by the searcher, the information becomes knowledge. The various relations between these concepts are as follows:

- **Data**: The raw material used as a vessel to store information in a given storage format (data structure).
- **Information**: The presentation of knowledge-concepts (usually through natural language) in a way that should enable knowledge gain. The latter can be achieved by the consumption of the information.
- **Knowledge**: A searcher's actual awareness and understanding of information, such that it can be applied when appropriate. Knowledge representation can be typed as either declarative (facts) or procedural (routines or sequences to perform a specific task).
- **Wisdom**: Some would argue that knowledge is superseded by wisdom, which has the added effect of being able to extrapolate and infer new knowledge from existing knowledge, through logical reasoning.

For example, a spreadsheet can contain various financial or other numeric data. Once (part of) the data is represented in a way such as to be useful (e.g. the sum of assets or the operating results), it becomes information. Reading and interpreting the information allows a searcher to become aware of it (knowledge), and act according to its implications (wisdom).

This classification strengthens the discerning of the following forms of retrieval [BAEZA-YATES1999]:

- **Information Retrieval**: In Information Retrieval, the searcher's information need is expressed as an *information request.* This request is often specified in the form of natural language, which is less structured and possibly ambiguous. As such, a document may not exactly be a complete match for the (vague) request, but may still be relevant. In Information Retrieval (IR) such a document would still be selected; based upon the request and the IS's interpretation of the documents, the relevance of each document is determined, which dictates whether or not a document is returned.
- **Data Retrieval**: In the case of data retrieval, the information need is expressed in the form of a *query*. The results are only deemed relevant when they exactly match the query's conditions. Obtaining a high precision is made possible due to the highly structured nature of the data, allowing for a specific instruction to only return the data containing information that's relevant for the information need.

Obviously, the difference in forms of retrieval leads to a significant dissimilarity between information requests and queries; formulating a query has a higher threshold than the formulation of a request, because the searcher needs to have extensive knowledge of (the internal structure of) the IS. Due to the fact that this research is not confined to dealing only with expert searchers (capable of formulating precise queries to reflect their information need), the focal point is on Information Retrieval.

3.2. Context of the Problem Space

With respect to the typical Information Retrieval (IR) process, the previously defined concepts relate to each other, the problem space, and the IS, in a certain way. We introduce the following schematic to illustrate their inherent relationships:

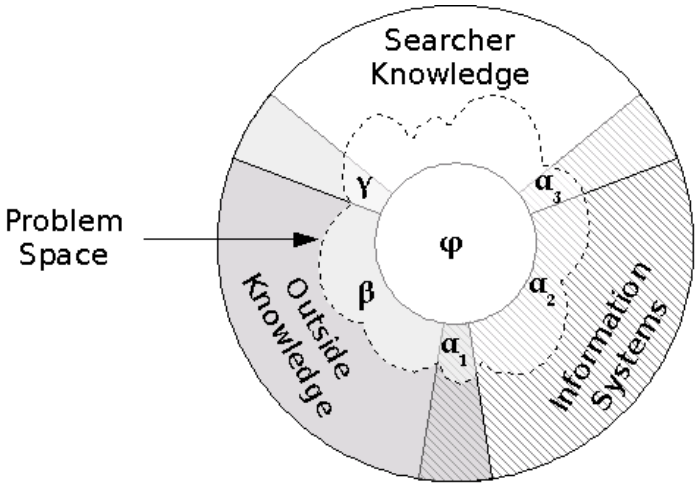

Illustration 1: Problem Space Context

- **Searcher Knowledge**: The total sum of knowledge that the searcher already possessed prior to issuing an information request on a given Information System (IS).
- **Information Systems**: The total collection of information within the reach of an IS. When executing a searcher's query, the result-set it yields will be a subset of this collection. Note that an Information System is defined as an agent, capable of providing information based on an information request. Consequently, the semantic scope of an IS in this research is more general than that of a *digital* Information Retrieval system.
- **Outside Knowledge**: The combined knowledge of all human beings, except the searcher. This could be interpreted as the 'society' the searcher is surrounded by, with respect to understanding and knowledge concepts.
- **Problem Space**: The information pertaining to the given problem space, that is required to overcome the related problem. The part of the problem space that falls outside of the searcher's knowledge is known as the *knowledge gap*. Note that some of the aspects of this area may still be unknown and/or unexplored. As such, part of this category may fall outside of the boundaries of all available information (as indicated by the problem space extending beyond all available information in the illustration).

As seen in the illustration, all of the above areas have some degree of overlap. The following overlapping areas are discernible:

- **Common Knowledge (γ)**:
  This area is defined as the intersection of searcher knowledge and outside knowledge. In contrast, the union of these areas defines all knowledge constructs that are present in the world.
- **Published Personal Knowledge (α_3)**:
  This area is defined as the intersection of searcher knowledge and information from the Information Systems. In contrast, the union of these areas represents the knowledge state a searcher can achieve by interacting with the Information System.

- **Published Outside Knowledge (α_1)**:
  This area is defined as the intersection between outside knowledge and (the knowledge published as information in) Information Systems. In contrast, the union of these areas represents all possible expansions of the searcher's knowledge, although outside knowledge constructs first need to be translated to information (and consequently, ported to an Information System), before the searcher can absorb them.

Although not apparent from the illustration, the intersection of searcher knowledge, outside knowledge and Information Systems represents general concepts, which are common knowledge *and* published as information in Information Systems. Due to the trivial nature of this category, it is omitted from the schematic.

The center of the problem space which does not intersect any of the 3 outer segments (φ), is the area of knowledge that is (as yet) unexplored and thus unavailable. An example of this type of knowledge is found in absolute facts, such as the relativity theory: although this theory was once unknown to mankind, the underlying effects have always been apparent in the universe. Once the theory describing this phenomenon was conceived, this absolute fact was moved from the area of unexplored area (φ) to the outside- or common knowledge. In relation to this illustration, wisdom can thus be perceived as a function to use existing knowledge to create more (inferred) knowledge, thereby exploring this area of unknown knowledge, and confining it even more. However, one could argue that new insights can also possibly lead to more questions, thereby even deepening the area of unexplored knowledge.

It is important to note that the above illustration is to be considered disproportionate and unspecific, because the ratio between the areas is ill-defined and subject to several parameters. For example, some situations might offer a relatively small and well-explored problem space, with all possible solutions to that problem being known to man. In this case, there is no unexplored knowledge (φ) regarding the Problem Space. Similarly, the contribution of an IS in terms of conveying information about concepts present in outside knowledge, could likely be higher than that of conveying information not present in outside knowledge (i.e. knowledge concepts being dormant in society, yet preserved in overlooked information).

The typical course of events leading to search behaviour, involves a person acknowledging his knowledge gap. The knowledge gap is the part of the problem space that is not fully present in the searcher's combined knowledge of the problem space. The person might experience contradicting thoughts and beliefs (a need for confirmation), or incomplete thoughts and beliefs (a need for clarification). In order to address this knowledge gap, the person defines the related problem space to the best of his knowledge: this problem space is described in terms familiar to the person. Then, the person establishes an approach to resolve his knowledge gap in the problem space and exhibits search behaviour (see *chapter 4: Search Behaviour*). Resulting from the search behaviour, information will be presented to the searcher, who then analyses and incorporates the relevant information into his personal knowledge. Depending on both the significance of the knowledge gap for the problem space and the complexity of the problem space, the searcher will at some point determine he's obtained sufficient knowledge to resolve his problem. If the searcher was unable to acquire the knowledge he needs, he might choose to continue the search, or abandon it altogether.

3.3. Information Relevance

There are various perspectives on which to judge the relevance of given information. The following paragraphs will explore two of those aspects:

3.3.1. Personal and Semantic Relevance

Because searchers by definition can not assimilate knowledge directly, they need to obtain information instead. Therefore, the searcher derives an information need from his knowledge gap, which he then needs to specify to an IS as a request. The resulting documents of this request will then fall in one of the following categories:

- **Semantic Relevance***:* When issuing a request to a given IS, the resulting set of relevant documents will fall within area α ($\alpha_1 \cup \alpha_2 \cup \alpha_3$). Documents that pertain to this area are known as the *semantically relevant documents* [STOJANOVIC2005], that the IS is capable of offering with respect to the given problem space. These are the documents that are relevant with respect to the searcher's request. However, part of this area (denoted as α_3) falls outside of the searcher's sphere of interest, since the concepts from this information are already known to him (to some degree).
- **Personal Relevance***:* Documents pertaining to area ($\alpha_1 \cup \alpha_2$) are known as *personally relevant documents* [STOJANOVIC2005], that the IS is capable of offering and that the searcher still needs, in order to better comprehend the given problem space. This area defines the searcher's actual information need that can be fulfilled by the IS: the knowledge constructs from within the outside knowledge are unavailable to the IS, and the exclusion of area α_3 corresponds with the fact that the searcher already possesses the knowledge constructs contained therein.

Most commonly, Information Systems are designed to yield the semantically relevant documents; documents that match the query in terms of the system's repository and domain model, rather than only those those documents that are personally relevant to the searcher's information need. As such, the searcher might be disappointed, as those results may have a big overlap with his present knowledge (α_3).

3.3.2. Relativity And Satisfaction Relevance

Following the issuing of the formulated information need, the IS will typically evaluate the input, and present the searcher with a number of (presumed) relevant resources. Consider a collection of documents A, B, C, D, E and F, given a specific problem space. The picture below represents the overlap and relation between documents and problem space:

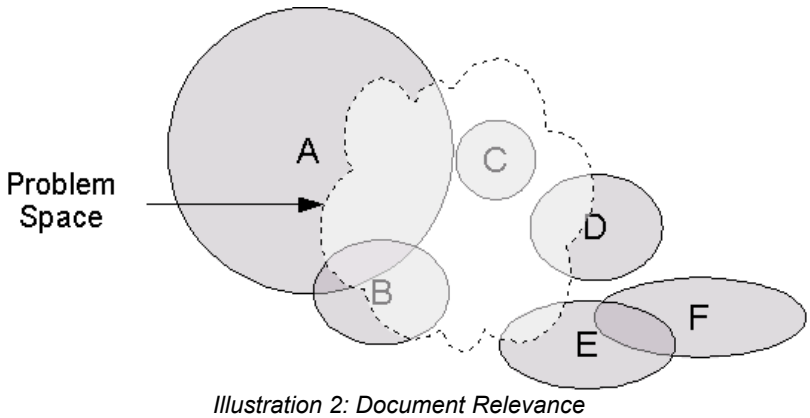

Illustration 2: Document Relevance

In light of this schematic, we can propose that there are two different measures of document relevance:

- **Satisfaction**: Documents are ranked with respect to the part of the problem space they affect. In this case, those documents that satisfy the largest part of the searcher's knowledge gap, are considered to be most relevant. This perspective on relevance focusses on the effectiveness of a given document in fulfilling the searcher's information need. In the above illustration, the ranking would be: A, B, C, D, E.
- **Relativity**: Documents are ranked with respect to the ratio of the part of the document that intersects with the problem space, to the part that doesn't. The documents of which most of their content falls within the problem space, are considered to be most relevant. In other words, this measure of relevance focusses on the efficiency of the document's ability to fulfil the searcher's information need. In the above illustration, the ranking would be: C, B, D, A, E.

In relation to *Illustration 1: Problem Space Context*, obtaining new knowledge from documents causes the concepts therein to shift from one part of the classification (area α_1 U α_2) to the published personal knowledge area (α_3). This expansion of the latter area entails the searcher gaining knowledge, at the expense of the IS' population of searcher-relevant information: information already conceptualised in the searcher's knowledge is henceforth deemed irrelevant.

Thus, exhaustive knowledge acquisition in terms of a given IS, relates to expanding the personal knowledge to the extent of the IS being void of relevant resources to expand this area. At this point the searcher would have extensive knowledge of this IS's repository and may choose to consult another IS if the information need is not yet satisfied.

3.4. Information Retrieval Process

Generally speaking, Information Retrieval is performed by a searcher, interacting with an IS, in the searcher's quest to find information that will fill his knowledge gap. When further dissecting the various elements of an IS, the following schematic representation of the IR process can be obtained:

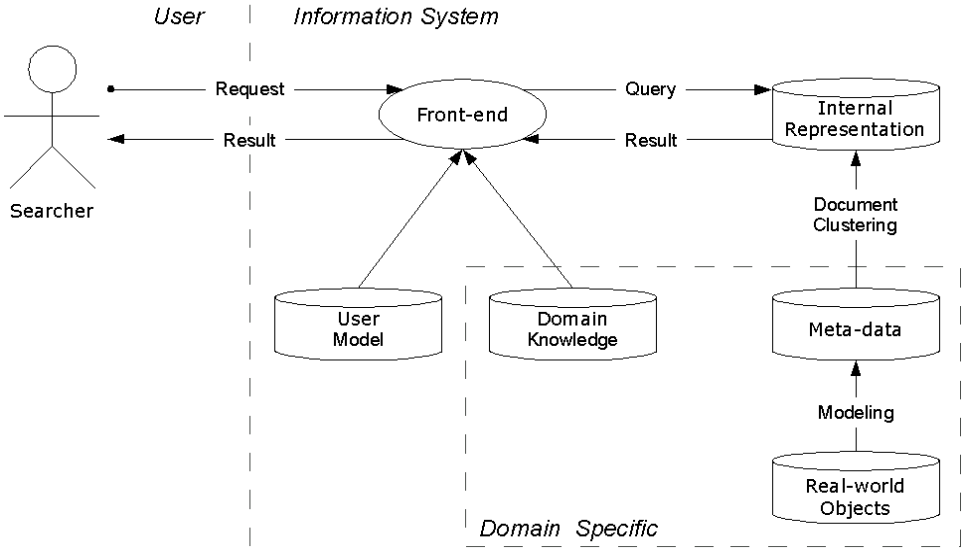

Illustration 3: The Information Retrieval Process

In the illustration above, interaction occurs between the searcher and the IS' front-end (a form of information retrieval), which interprets the searcher's information request and executes a query (a form of data retrieval) on the internal representation of the meta-data. The query's results can then be analysed and used to help satisfy the searcher's information need. Also, in accordance with developments in the field of User Modelling, the front-end may be able to make use of a user-model (involving various details of the searcher's background) and other domain knowledge (i.e. business rules applicable to the given domain) to further narrow the set of results that is offered to the searcher.

The internal representation can be arrived at through document clustering, by making use of techniques such as Latent Semantic Indexing, Self-Organizing Maps or other techniques. This however falls outside the scope of this research.

4. Search Behaviour

"Seek, and ye shall find." - Bible, Matthew 7:7

This chapter serves to make an analysis of the aspects that affect the searcher's state during the various stages of the exploratory search process, and how this expresses itself in actual search behaviour.

4.1. Exploratory Searching

Exploratory searching [WHITE2005] is a search process, in which the searcher is (initially) unaware of the exact nature of his information need and does not have a defined target. This implies that the searcher is then also unable to express or formulate this need, due to a lack of knowledge of the problem space. More explicitly, during such a process the searcher's state is not constant, but changes continuously, as the searcher gains a increasingly better understanding of his knowledge gap. Several of these states, all of which refer to a searcher's *information need level* have been named by [TAYLOR1968] which are defined as follows:

- **Visceral Need**: A vague sense of dissatisfaction, due to an actual, but unexpressed need for information.
- **Conscious Need**: A conscious mental description of the need can be made, in the form of a (possibly ambiguous) narrative or as examples/analogies.
- **Formalized Need**: The need can be formalized as a rational question-statement or a topic with clear and concise boundaries.
- **Compromised Need**: The representation of the need within the constraints of the system and its information repository.

When looking at it from the perspective of fuzzy logic, the information need awareness can be considered a linguistic variable, which can take the value of any of the information need level denominations mentioned above.

As the searcher consumes information, his knowledge of the problem space increases. This also causes a shift from the searcher's anomalous state of knowledge ([BELKIN2005], cf. knowledge gap). With each consecutive need level, the searcher's ability to express his information need (and the corresponding accuracy) is improved. Note that a search process need not necessarily traverse each of these stages; a searcher might either not be sufficiently motivated to continue searching or might have fulfilled his information need in one of the earlier information need levels through serendipitous discovery (i.e. stumbling upon the required relevant information by chance).

Several of the *information seeking strategies* (ISS) from [BELKIN1993] are able to represent the various information need levels. Each strategy is represented as a set of values within the following set of dimensions:

- **Goal of Interaction**: The goal of the interaction might be to *learn* about (characteristics of) an object, or to *select* useful objects for retrieval.
- **Method of Interaction**: This dimension pertains to either *searching* (looking to find a known object) or *scanning* (looking for something interesting in a collection).
- **Mode of Retrieval**: The mode of Retrieval can either be by *recognition* (identifying relevant items while stimulated by the possible options), or by *specification* (looking for identified objects).
- **Resource Considered**: Finally, this dimension describes the resource focus, being either the actual object (*information*), or information about this object (*meta-information*).

In [BELKIN1994] the set of dimensional values [Learn, Scan, Recognize, Meta-information] is described as "*a situation in which a person needs to learn about characteristics of the knowledge resource before the information search can begin. This can also be understood as the ISS associated with an unformulated and unspecified information problem.*"

This accurately represents the visceral need level. Also, the set [Select, Search, Specify, Information] is described as "*a prototypical example of a well understood information problem, in which the goal of the interaction is not to learn about the system, but to select items which can be specified by the user.*", which depicts a formalized/compromised need level.

Similarly, we can use the information seeking strategies to differentiate the various aspects of each of the other information need levels:

Need Level	*Goal*	*Interaction*	*Retrieval*	*Resource*
Visceral	Learn	Scan	Recognize	Meta-information
Conscious	Learn/Select	Scan/Search	Recognize	Information
Formalized	Select	Search	Specify	Information
Compromised	Select	Search	Specify	Information

4.2. System Awareness

Note that the aspects of the formalized and compromised levels only differ in the form in which the need for information is communicated. In the compromised need level, the information need is expressed in terms of the IS; the searcher understands the language used in interacting with the IS, as well as the structure of the documents within the IS. The compromised need level, therefore, implies a greater familiarity with the IS' interface and the available collection of documents.

Because of this difference, we propose that the compromised need is not an awareness level, but rather a state of being able to translate the high information need awareness (formalized) to an IS.

Due to this clear distinction between a searcher's information need awareness and his awareness of the system, we introduce an additional dimension, called *system awareness*. This dimension positions the aforementioned information need levels as follows:

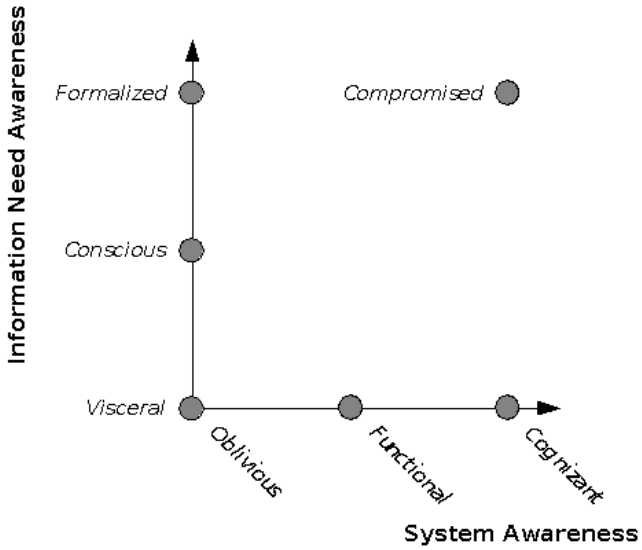

Illustration 4: Information Need Awareness vs. System Awareness

Much like the various levels of information need awareness, system awareness can also be considered as a linguistic variable, the value of which can range through the following denominations:

- **Oblivious**: Initially, the searcher lacks any knowledge of both the (structure of the) repository that the IS can deliver and the language used to interact with the IS.
- **Functional**: Once the searcher has gained some understanding of the language used by the IS, the searcher is increasingly proficient at expressing his perceived information need to the IS.
- **Cognizant**: In this level of system awareness, the searcher is 'fully informed': this implies that the searcher has a sufficient system awareness to formulate his perceived information need as a IS-specific query.

Note that by this definition, it is prerequisite to have a thorough insight into the language used by the IS, prior to gaining sufficient understanding of the IS's repository.

From this we can infer that the search process of a given searcher can be described as a path between these two awareness dimensions; such a development path can be formed with respect to the progression of the searcher's information need awareness and system awareness. These *Awareness Development Paths* (ADP's) will vary from person to person. We will illustrate this concept through several exemplary ADP's, numbered 1 through 5 in the following illustration, showing the development of the searcher's awareness dimensions during the interaction with a given IS:

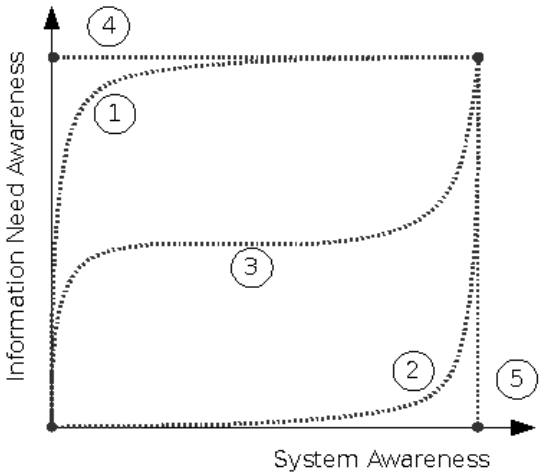

Illustration 5: Exemplary Awareness Development Paths

In the above illustration, the following ADP's are identified:

- **ADP1 - Taylor-esque**: This pattern adheres to Taylor's classic progression of need levels: the initial part of the interaction between the searcher and the IS is focussed on increasing the searcher's awareness of his actual information need, followed by the increasingly accurate communication of this need in terms of the IS. This path is typical for people who are unable to express their information need (and thus have a low information need awareness) and strive to use the information system to gather more information in general areas of interest. Once they become more aware of the exact nature of their information need, they will formalize a specific request to the system. By doing so, they will learn more about the system (and increase their system awareness) and consequently be more proficient in expressing their information need in terms of the IS.
- **ADP2 – Experimental Search**: Once the decision for a particular IS has been made, some people will aim to gain system knowledge prior to increasing their information need awareness. In effect, they are postponing the resolution of their information need, in favor of experiment with the IS, to increase their understanding thereof.
- **ADP3 – Arrested Development**: Quite often, people will fail to attain a sufficiently high degree of information need awareness and are therefore unable to express their need satisfactorily. As a result, the progression of their information need awareness will stagnate. When this occurs, people tend to attempt to formulate *tentative queries* [WHITE2005] and present these to the system and, by doing so, gain increased system awareness. This will continue until the searcher experiences a breakthrough in his information need awareness, which is enabled by one of the results from the tentative query.
- **ADP4 – Formal Searcher**: For a given information need, people will always traverse the lower stages of information need awareness (e.g. visceral and conscious). However, this need not necessarily be done in unison with an IS; a searcher may have been able to acquire an accurate perception of his actual information need, even prior to using an IS. As such, with respect to the IS, the searcher is already in the formalized need level and will use the IS to (learn to) express his need in terms of the IS.

- **ADP5 – Skilled Searcher**: Conversely, people may have a high degree of system awareness, yet fail to understand the exact nature of their information need. This can occur when the searcher is in fact an expert of the domain in question, but is unable to make a choice in what would best suit his needs. This can occur when the searcher is confronted with too many options (as a form of information paralysis) and requires additional help to make a decision.

The set of awareness development paths is practically endless, as a given searcher can start from any arbitrary point in the spectrum of information need awareness and system awareness, and proceed from there. The path between these dimensions indicates a progression of the searcher's awareness, as well as time.

4.3. Search Strategies

The searcher's degree of information need awareness calls for a specific search strategy, that the searcher will employ to further increase his information need awareness. Various models offer different types of search strategies (such as [WEICHOO2000], [SONNENWALD2001] and [KUHLTHAU1993]), but fail to connect these to either the searcher's information need awareness or related information need level. For each of these models, the various search strategies relate to the information need levels as follows:

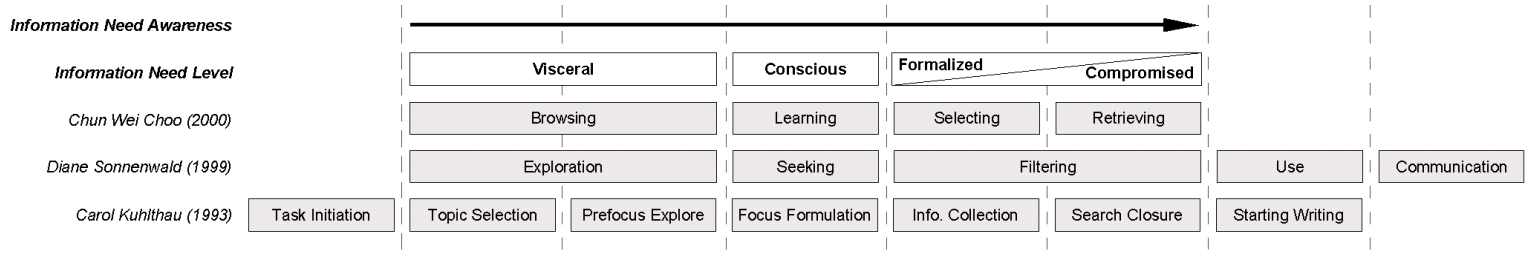

Illustration 6: Search Strategies Related To The Information Need Levels

Note that some of these information strategies aren't part of the user's search behaviour, as they precede (*Task Initiation*) or succeed (*Use, Starting Writing* and *Communication*) the search for information. As such, they are not considered part of search behaviour, but rather as information behaviour.

From all of these behaviour models, one unified model emerges, offering 4 general strategies that each confers with one of Taylor's need levels. Although these strategies vary in name for each of the aforementioned models, they each describe the same aspect of the user's search behaviour. For the unified strategies, we will reuse the terminology used by Choo, for the close resemblance with the terminology as used in this research:

- **Browsing**: The searcher has general areas of interest, but is unable to describe them. In this strategy the searcher attempts to improve his perception of his information need, through negotiation with the IS. This is accomplished via recognition of relevant information and its underlying characteristics; more specifically, by scanning a broad range of sources and judging them based on recognizing terms and concepts from the easily accessible (meta-)information.
- **Learning**: The searcher aims to further increase his knowledge of the problem space, by actively searching for topics of interest, within the general areas of interest that resulted from the previous strategy (Browsing). His understanding of the extent of the problem space has somewhat increased, but the information need awareness is still somewhat blurred. This results in a broad and unrefined specification of his information need.

- **Selecting**: Having increased knowledge of the problem space, the searcher is able to compose an accurate description of his information need. As such, the searcher will formalize a plan in the form of a formal description of the searcher's information request.
- **Retrieving**: The searcher's precise description of his information need is adapted to the information system, generally aiming for high precision, at the expense of recall. In this scenario, the personal relevance of the results will closely relate to the semantic relevance thereof.

4.4. Search Flow

Based on models from [MARCHIONINI1995], [TAYLOR1968], [WILSON1997], [BRUCE2005], [NIEDZWIEDZKA2003], a unified flow chart modelling the search process is derived:

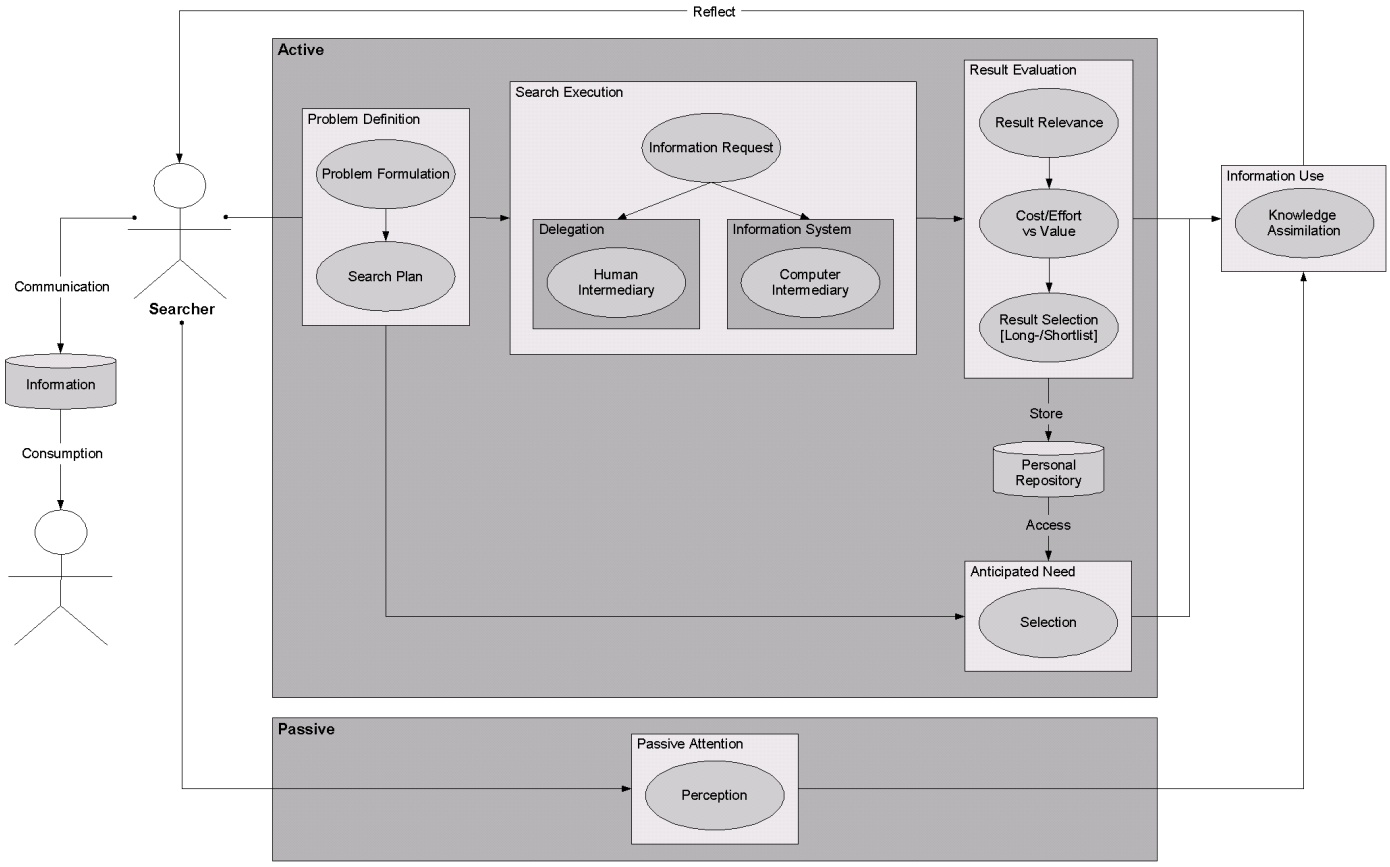

Illustration 7: The Search Process, As A Flow Of Operations

The unified search flow chart encompasses suggestions made by the individual contributors, the most significant segments of which are as follows:

- **Active Search**: A searcher who experiences a need for information will initially try to define the problem he faces. This is done by composing some type of plan to fulfil his knowledge gap and consequently trying to express or formulate his problem, to be used for the information request. The actual search execution can take two forms ([NIEDZWIEDZKA2003] and [TAYLOR1968]:

- **Information System**: By conveying his information request to a digital IS, the searcher initiates the search process (as represented in illustration 3: *The Information Retrieval Process*) and acquires the results of his request.
- **Delegation**: Rather than engaging in the actual search himself, the searcher may delegate this task to someone else, by conveying the information request to a human intermediary. This intermediary will then act based upon this information request and (essentially) perform the entire search process recursively instead of the actual searcher. In this recursion, the intermediary will become a searcher himself, formulating his own search plan and executing it. The results thereof are then returned to the original searcher. Further insight into this process is provided in *chapter 5: Practical Intermediation Models*.

The result of the search execution is then evaluated; firstly on the relevance of the information and secondly on the cost or effort to consume the information. Ultimately, based upon these evaluations a selection of the information resources is made and consequently consumed.

- **Passive Attention**: In contrast to active search, *passive attention* [WILSON1997] is not initiated by the searcher. Although the searcher experiences (but not necessarily perceives) an actual knowledge gap, he is not actively searcher for information, but rather absorbed it passively (e.g. listening to the radio or from an advertisement).
- **Anticipated Need**: [BRUCE2005] describes the concept of *anticipated need*, in which a searcher encounters information resources that fulfils an information need he does not yet have. Instead of consuming them, the searcher can choose to *store* the resources (or references thereto) in a personal repository, based on the assumption that they will prove relevant later on. When the need (that the searcher anticipated he would have) emerges, the searcher is able to formulate a search plan involving *accessing* and selecting resources from his personal repository. An example of such a personal repository are bookmarks or favourites, in an internet browser.
- **Communication**: Once given information has been consumed (and has become knowledge) it can be communicated to someone else. To do so, it first has to be reformulated and expressed as information (e.g. writing, speech, etc.). This information can then be consumed by others.

5. Practical Intermediation Models

"Those that won't be counselled can't be helped."
- Benjamin Franklin

Besides individual (unassisted) search behaviour, people in need for information can be assisted during their search process. In everyday life, such scenario's are quite common, yet (digital) Information Retrieval often ignores this aspect of the search process. This chapter explores the use of intermediation strategies to support each of the searcher's search strategies for digital Information Retrieval, and seeks inspiration from (human-to-human) practical intermediaries that exist in everyday life.

5.1. Intermediated Information Systems

Most existing Information Systems only employ one form of interaction to interact with the searcher, regardless of the searcher's information need awareness. Given that the searcher's degree of information need awareness will result in different search strategies (see *paragraph 4.3: Search Strategies*), this would imply that the support that the IS offers to the searcher is not always optimal and therefore subject to improvement. Improvement can be made by employing different intermediation strategies for the different stages of a searcher's information need awareness and system awareness, allowing the intermediary to accurately address the searcher's current state in his Awareness Development Path. A more detailed conceptualisation of the various intermediation strategies between the searcher and the intermediary can thus be illustrated as follows:

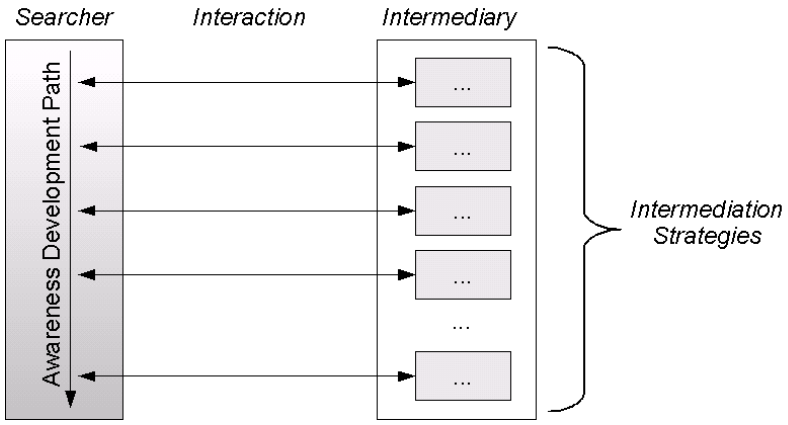

Illustration 8: Consecutive Intermediation Strategies During Development Of Awareness

Note that there can be more intermediation strategies than the information need levels as defined by Taylor, since these levels are not formal categories, but actually shade into each other. In the above scenario, each intermediation strategy continues until the searcher has reached a higher information need- or system-awareness. For example, a searcher who's operating in a visceral level might benefit from an intermediation strategy aimed at brainstorming [VANVLIET2007], whereas the same strategy is considered too cumbersome for a formalized need level, which in turn would benefit more from intermediation aimed at providing structure and abstraction of the concepts in the problem domain.

This implies that the system must have a means to detect what information need awareness the searcher is currently affected by, in order to know when to apply a different intermediation strategy. Rather than asking and interrupting the searcher (who might not even know the answer himself), this information should be derived from the searcher's actions and behaviour. Because the searcher's need level is dynamic and "*these four levels of question formulation shade into one another along the question spectrum*" [TAYLOR1968], detecting the searcher's information need level is not (or even vaguely) formally described.

Depending on the information need awareness, people may employ different strategies to search for information (see *paragraph 4.3: Search Strategies*). In real-life, there are several scenario's in which people find themselves in need for information and, due to the uncertain nature of their need, will consult an intermediary to aid in their quest for information. By studying such scenario's, in which people engage in exploratory searching, we are able to extract successful intermediation models for assisting and guiding a searching to obtain a higher information need. In the following paragraphs, several intermediation models employed in practice, by real-life intermediaries, are described. Each of those models consists of various intermediation strategies, that apply to the varying stages of the searcher's information need awareness.

5.2. Physicians

Patients that have a physical ailment, will most often be unaware to the exact condition that affects them and will only be able to describe its symptoms. Much like exploratory searching in IR, patients may seek the help of a physician to aid them in discovering what is affecting them (their knowledge gap) and apply the appropriate treatment (the applicable relevant documents). From the model of [LARSEN1997] the following intermediation strategies that are employed by physicians to treat patients are derived:

- **Observation and Interpretation**: In this stage, the patient is communicating his awareness of the symptoms to the physician. The physician in turn listens to and observes the patient, to start establishing a mental model of the patient's condition. Finally, the physician summarizes the patient's story, to make sure his message was communicated successfully. This stage results in the patient feeling understood (reducing anxiety), while providing the doctor with a general sense of the problem context.
- **Interview and Examination**: The physician tries to clarify the problem context, by linking this general view to his intrinsic medical knowledge. For example, the patient could only be aware of a subset of all his symptoms, so the physician will interview and examine the patient, to find and/or rule out possible other symptoms. Once the physician has gathered enough information pertaining the condition, he will match the problem context to medical knowledge (either already available, or to be obtained by sessions of information seeking), and come up with a diagnosis.
- **Devise Treatment**: The diagnose links the found problem context to comparable problem situation(s), which have been established through medical research. This research might have devised several relevant treatments for the problem condition, some of which may be inappropriate in the patient's personalized context (due to allergies, medical history, etc.). The physician comes up with a treatment, compatible with the patient's situation.
- **Apply treatment**: The treatment is conveyed to the patient, and applied to remedy the patient's ailment. The application can occur under the physician's supervision, or at a scheduled time by the patient himself.
- **Treatment Monitoring**: A follow-up appointment will be scheduled, at which time the physician can determine the effect of the treatment on the problem, and adjust the treatment accordingly.

5.3. Salesmen

It is the task of the salesman to aid customers, that are in need of a certain product to fulfil a certain task, in discovering what product best suits their needs. Applying the analogy to Information Retrieval, the salesman will interact with the customer to discover the exact nature of their need (their information need) and consequently recommend applicable products (relevant documents) to satisfy the need. The models of [BERGMANN2002] and [AGMRC2007] enable the translation of the analogy of the salesman to the field of Information Retrieval, by introducing a virtual sales assistant that uses his knowledge of products *and* of clients to recommend suitable products. From this model we can derive the following consecutive intermediation strategies that a (virtual) salesman would employ:

- **Experience**: The salesman needs a comprehensive understanding of the products in his assortment and the features that each offers. Also, the salesman needs to be able to relate a customer's need to the applicable features. This understanding can come from studying the assortment, as well as experience with customer interaction.
- **Initial Contact**: The salesman needs to initiate contact with the customer and offer to help him achieve his goal (i.e. fulfil his need).
- **Requirements Acquisition**: The salesman interviews the customer and attempts to ascertain the nature of their need. Obviously, the customer will not wish to answer an endless list of questions. This implies that the questions must be as discriminative as possible, in order to effectively narrow down the set of applicable products. In order to do this, the salesman will need to use his knowledge of the store's assortment to generate relevant questions. Also, the salesman will need to be able to translate his knowledge of the assortment to terms and concepts that the customer will understand, as the customer need not necessarily be an expert in the domain of such products. Generally, there are two tactics for capturing and formalising the customers information need [SHIMAZU2002]:
    - **Navigation-by-Asking**: The salesman generates a dialog, based upon the features of the products in the available assortment. The customer is asked to choose between features that interest him, thereby navigating through and reducing the set of features of the assortment.
    - **Navigation-by-Proposing**: Based upon some initial information derived from the customer, the salesman is able to propose products to the customer. Based upon this proposal the customer is able to navigate through the assortment by evaluation the proposing product with respect to his need. The customer can, for example, explain that he wants something cheaper or more powerful. Based upon these explanations, the salesman is then able redefine his perception of the customers information need and make additional proposals.
- **Product Search**: Based upon the salesman's perception of the customer's need, he can evaluate the alternatives from the assortment and create a consideration set. Generally, there are two tactics to evaluate the applicability of given products:

- **Content-based Approach**: Based on the extracted features from the requirements acquisition, the sales man is able to determine the similarity between these extracted features and the products from the assortment. The products that exhibit the greatest similarity can be included in the consideration set.
- **Collaborative Approach**: This tactic emerges from "*the potential for a collaborative approach to product retrieval where users are matched to products based on past behaviour or consumption history*" [BERGMANN2002]. As such, the salesman's experience with other customers, who exhibited a similar need, can be used to elect products into the consideration set. This similarity can be determined by looking for correlations between customers in terms of their ratings assigned to items in a user-model.

- **Product Presentation**: At this point the products from the consideration set are presented to the customer. The type of product presentation can vary upon the customers individual information need as well as the type of the product (text, sound, video, etc.). Also, "*the presentation form of an information entity should be chosen so that the respective message is communicated with a minimum effort of the user*" [BERGMANN2002]. As such, the customer should be enabled to easily browse through the products, review them, and compare two or more different products.

  Note that the requirements acquisition and the product presentation are repeated in a cycle, which may be repeated several times before a satisfactory product is found. As such, "*requirements acquisition and product presentation often cannot be completely separated*" [BERGMANN2002]. This implies that, even during the product presentation, the customer is able to further refine the set of products and provide additional feedback on a given product, resulting in a new product search and consideration set.
- **Sale**: When one of the products from the consideration set is accepted by the buyer, the sale is confirmed. The customer supplies the required (economic) cost and the seller delivers the product in question.
- **Evaluation** (optional): The salesman can check to verify the customer's satisfaction with the delivered product. If needed, the product can be revised, based on the customer's feedback. Additionally, this step provides reflection (for future deals) for both salesman and customer.

5.4. Librarians

In [TAYLOR1968], Taylor describes the '*communication between inquirer and librarian during negotiation process*'. Five stages (know as "*filters*") are identified, that the librarian uses in his intermediation task:

- **Subject Definition**: By defining the subject, the librarian can more accurately delineate his search space, in order to optimize his scope for the retrieval of (possibly) relevant answers. Once the subject's scope is relatively clear, the librarian might be inclined to '*make a brief search to determine the extent of the subject*', to determine the extent of the subject, based on the information available to him.
- **Objective and motivation**: Understanding the motivation behind the searcher's request for information also helps the librarian in verifying and adjusting the established subject definition. Additionally, the insight in the searcher's point of view influences the '*shape, size and form of possible answers*'. Finally, searchers often '*cannot define <u>what</u> they want, but they can discuss <u>why</u> they need it*', making the motivation behind the search the only concrete source of direction for the librarian.

- **Personal Background of inquirer**: Knowledge about the searcher's context in terms of intrinsic knowledge, awareness and familiarity with the library has relevance to the negotiation process. It for example determines the urgency, dialogue level and critical acceptance of results. Having interacted with the searcher before provides some inside knowledge that will ease the dialogue.
- **Relationship of inquiry to file organization**: The librarian "*becomes a translator, interpreting and restructuring the inquiry so that it fits the files as they are organized in his library*". It is up to him to transform the information request into a system query to allow for an efficient system search, yielding the right answers for the searcher. Aside from using the library system, the librarian might be inclined to use (refer to) additional sources of information, e.g. other librarians, or recently used information.
- **Provide acceptable answers**: The librarian must address the searchers' information need by providing information that is correct (relevant) for the searcher. The searcher has a certain expectation of acceptable answers in terms of format, size, completeness, soundness, etc. The librarian must provide the right amount of information in the right format to address the searcher's information need.

5.5. Evaluation

It is evident that the salesman model and the physician model have striking similarities. Even though the salesman model explicitly mentions 'Experience' as a strategy in which information is gathered about both products and customer interaction, the same principle holds true for the physician, who has spent time studying medicine and gained practical experience during internships. Aside from this dissimilarity, both models are evidently equivalent.

However, the librarian model would (at first glance) not seem to exactly correspond to the previous models. The reason behind this is the fact that the librarian model is not composed of consecutive steps, but rather a set of 'filters' as arbitrarily applied by librarians in the intermediation process. When examined closer, these filters are translatable to both physician and salesman model, yet not necessarily in a direct 1:1 relationship. The following table relates each of the strategies from the various models and offers the underlying description for all of the individual models, to form a series of generic strategies:

Physician	*Salesman*	*Librarian*	*Strategies*
	• Experience		Acquire Intermediation Information
• Observation and Interpretation • Interview and Examination	• Initial Contact • Requirements Acquisition	• Subject Definition • Objective and Motivation • Personal Background of Inquirer	Requirements Acquisition
• Devise Treatment	• Product Search	• Relationship of inquiry to file organization	Convert Request to Query
• Apply Treatment • Treatment Monitoring	• Product Presentation • Sale • Evaluation	• Provide Acceptable Answers	Result Presentation

With respect to the field of (digital) Information Retrieval, these generic strategies can be described as follows:

- **Acquire Intermediation Information**: During this strategy, that precedes the actual intermediation process, the intermediary is trained to gain the information necessary to perform the required intermediation tasks in the following strategies.
- **Requirements Acquisition**: In order to acquire an accurate perspective of the searcher's information need, the searcher is enabled to express his request in his own terms (navigation-by-asking). The intermediary then attempts to help the searcher increase his information need awareness, by offering related terms and keywords that may also be relevant to his information need. This helps the searcher narrow down the exact nature of his information need.
- **Convert Request to Query**: Once the intermediary understands the searcher's information need, the intermediary is able to translate the searcher's request to the indexes of the documents. This translation is aimed at increasing the searcher's system awareness. Using this translation of the searcher's request, the intermediary is able to execute a query on the set of documents in the IS's repository.
- **Result Presentation**: The results of this query are then presented to the searcher. By using the searcher's feedback on given documents (navigation-by-proposition) the query can be further refined and recast to the IS's repository, in order to offer an increasingly relevant set of documents.

The primary difference between an intermediary and a traditional IS, lies in the intermediary's knowledge of searchers and their perception of documents. A traditional IS knows a great deal of the documents it offers, but is less proficient of linking the searcher's expression of his information need to (the indexes of) the documents. As a result, traditional IS's require searchers to express their need (i.e. their request for information) in the indexes that the IS uses to index its documents. These indexes need not correspond to the searcher's perception of those documents, causing problems with the interpretations of the searcher's information need. Intermediaries, however, use their knowledge of the human perspective on documents to allow searchers to express their need in their own terms, and (transparently) convert this to the system's indexes.

6. Consolidated Intermediation Model

"Quidquid latine dictum sit, altum sonatur."
- Origin unknown

From the practical intermediation models and the observations made from them, we can now derive a new model that consolidates the practical strategies, for the purpose of supporting (digital) exploratory search. Generally speaking, this involves gaining knowledge about both documents, the searcher's request and a way of fulfilling this request with the aforementioned documents. The following paragraphs in this chapter will each describe a part of the resulting consolidated intermediation model:

6.1. Acquire Intermediation Information

This part of the consolidated model focuses on acquiring the information necessary to offer intermediated support for the searcher.

6.1.1. Intermediation Information

Generally, there are two ways for a searcher to express his information need; in terms that a consumer would express his need and in terms that a producer would describe his product. For example, a searcher in need for a safe car, might express his need in terms of "safety" or "security", whereas an applicable product might be described using the related (but non-equivalent) terms "airbag" or "brake-assist". As such, the indices that apply to the documents in the IS's repository (internal indices) don't always match the indices that a searcher might use to express his need (external indices).

From the practical intermediation models, we can see that the intermediaries use several sources of information during the intermediation process. The first is derived throughout the intermediation process, whereas the others are obtained through training and/or experience:

1. **User Information**: The information that the intermediary derived from the searcher during the interview/negotiation process, which can be used to shape the intermediary's perception of the searcher's actual information need. The user information can be determined through requirements acquisition, which will be detailed in the next paragraph.
2. **Meta-Information**: The intermediary's knowledge of the intrinsic characteristics (internal indices) of the documents in the IS's repository. Knowledge of this type of information by the searcher, results in a higher system awareness (see *paragraph 4.3: Search Strategies*).
3. **Human Perception Information**: Using the intermediary's knowledge of searchers, and (specifically) the way they view the applicable documents, he is able to translate the searcher's need (external indices) to relevant documents' characteristics (internal indices).

Typically, search-engines require searchers to translate their information need to internal indices, even though this translation may be biased or incomplete. Even though in some cases (and trivial, non-ambiguous terminology) the internal indexes seem to match the specified keywords due to both parties sharing natural language, the translation still needs to be made. This implies that such search-engines assume a higher system awareness of the searcher, even though this awareness may be too low. An IS that uses human perception information to facilitate the translation from external to internal indices, would most likely yield results that are a better match to the searcher's expressed information need, because the external indices would better correspond to his vocabulary.

6.1.2. Human Computation

Meta-information can be acquired in the same ways as existing search-engines do, through document clustering. Human perception information, however, cannot be acquired in the same manner, as it implies a human perception on the (external) indices that are applicable to documents. For computers, this is obviously an open AI problem. A possible solution to this is offered as *distributed knowledge acquisition* [VONAHN2004]. Applying this concept onto information retrieval would call for groups of people to (indirectly) aid other searchers; people could aid searchers by applying indices which they perceive to be relevant to given documents. People could be convinced to index and categorize objects from a given domain through the use of *games with a purpose* [VONAHN2006]; whilst playing such a game, people will (implicitly) acquire meaningful data over specific objects. This data could then be used to relate the searcher's external indices to actual documents. Employing humans in this fashion, in order to solve open AI problems, is called *human computation* [VONAHN2006].

In such a *game-with-a-purpose*, users (in a multiplayer game) could be confronted with an document and be asked to guess what the other players might be thinking. When the players have nothing in common, except from the document in question, the players will ultimately attempt to provide terms or keywords that best describe the object. Using techniques such as taboo words (words that are disallowed in the game), it can be ensured that the set of keywords for a given document are as broad and as complete as possible. The more people that play the game (and thus provide input for a given document), the more accurate and error-free the set of keywords would be; a given incorrect keyword is less likely to be submitted that a correct keyword, meaning that incorrect input can be eliminated through statistical analysis. For more measures to assure the quality of the keywords, see [VONAHN2004]. The acquisition of external and internal indices, through human computation and document clustering respectively, can be illustrated as follows:

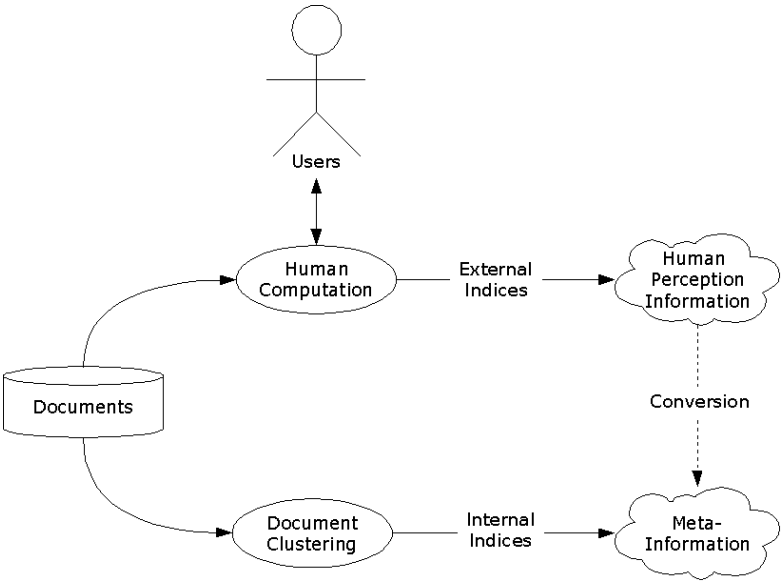

Illustration 9: Acquisition Of Intermediation Information

Human computation can thus be seen as a parallel process, that can run alongside the intermediation process. Although a certain minimum of information must have been acquired in order to perform intermediation in meaningful way, both human computation and intermediation can run independently of one and other.

6.1.3. Relationship-map

Once external indices have been acquired on a given document (through human computation), these indices need to be structured in some form of hierarchy. Such a hierarchy of indices should not separate the set of characteristics into a strict categorical directory listing, as it may well result in "*arbitrary hierarchical arrangements*" and would suffer from "*subjectivity of rating and annotating resou*rces" [WEBLIMINAL2006].

In order to perform the requirements acquisition (as described in the next paragraph), the IS will need to have an understanding of the various types of linguistic relationships that might exist between the external indices. A relationship-map should keep track of the various types of relations that might exist between the indices, such as holonyms / meronyms (*x part of y*) and hypernym / hyponym (*x is type of y*). Such a relationship-map could automatically be derived by applying the total set of indices to a lexical database, such as Princeton University's *WordNet* [WORDNET2007].

For example, in the human computation stage, users playing the *game-with-a-purpose* could be confronted with an "Airbus 360" (either in graphical form or textual description). The information acquired from the users would take the form of the external indices that most users believe to apply to the Airbus. These could include "wing", "airplane" or "vehicle". Using a lexical database, these (and other) indices could be structured as follows:

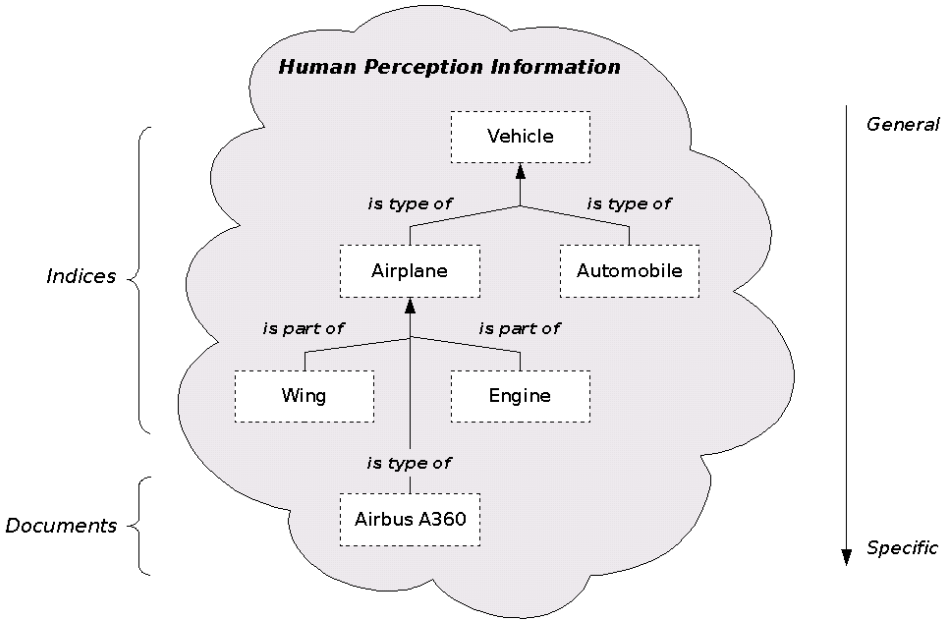

Illustration 10: Identifying Relationships In Human Perception Information

The structure derived from the lexical database can be used to help define the contribution (grain, focus, relation) of the external index to the actual object indexed.

6.2. Requirements Acquisition

The requirements acquisition stage serves to improve the searcher information need awareness, to the point where he is able to express his need (in external indices), as well as enabling the intermediary to create a perspective on the searcher's information need. From the practical intermediation models, we can derive that during the requirements acquisition stage the intermediary applies two separate strategies to improve his perspective of the searcher's information need. Note that both strategies revolves around the external indices, as the searcher may well be unaware of the applicable internal indices:

- **Passive Acquisition**: Initially, the intermediary enables the searcher to narrow down the area of his information need, by using the most discriminative indices as possible. By using the most discriminative indices, the intermediary needs fewer indices to be able to proceed to the following strategy. This matches Taylor's intuitive approach; "*We work from the general to the specific*" [TAYLOR1968].
- **Suggestive Acquisition**: Based upon the intermediary's perspective on the searcher's information need, additional suggestions for indices are offered that the intermediary assumes to be relevant. This perspective is based upon the indices that the searcher previously indicated to be relevant to his need. The suggested index will thus have some type of relationship with the original indices, that emerged from the relationship-map. The suggestions will serve to further narrow down the area of the searcher's information need by following the path through the relationship-map, from the initially relevant indices (general) to the more specific indices.

Passive acquisition can be accomplished by presenting the searcher with a set of (otherwise random) external indices. These indices would come from (or near) the top of the relationship-map, as those indices are most discriminative. The searcher would then be able to cast a *judgement* on those indices relevant to his information need. In case none of the indices from the set are relevant to the searcher, the set should automatically update to include other (equally random) indices.

The suggestions are based upon the intermediary's perspective on the searcher's information need, which in turn is based upon the searcher's judgements. Once it becomes clear that the suggestions receive positive judgements from the searcher, it can be assumed that the intermediary's perspective on the searcher's information need is correct. As discussed in [VANVLIET2007], the suggestions stop "*when the reaction of the searcher seems to be reasonably predictable, and suggestions for extensions are exhausted*". At which point, the interaction can proceed to the next strategy, as discussed in the following paragraph.

6.3. Convert Request To Query

To improve the searcher's system awareness, the intermediary's perception of the searcher's information need, expressed in (judgements on) external indices, needs to be translated into internal indices. This matches Taylor's description of a librarian, who "*becomes a translator, interpreting and restructuring the inquiry so it fits the files as they are organized in his library*" [TAYLOR1968]. In effect, this step interprets the searcher's request for information (expressed in judgements on proposed indices) into a query on the total set of documents.

This conversion process is accomplished by determining the correlation of the external indices to the internal indices on a given document, as defined in the human perception information and meta-information respectively. Schematically, this can be viewed as follows:

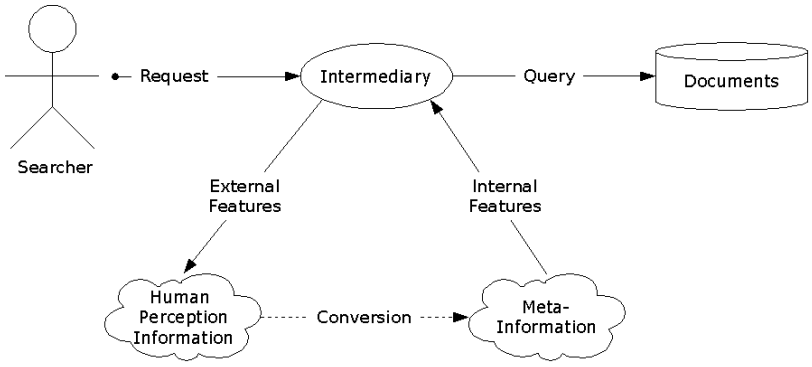

Illustration 11: Conversion Of Request To Query

6.4. Result Presentation

The result presentation stage serves to improve the searcher's system awareness, to the point where he is able to express his information need in internal indices, and consequently return the resulting documents that match those internal indices. Additionally, relevance feedback is employed to refine the query.

Once the query is executed and a result-set is obtained, the results can be presented to the searcher for evaluation and feedback. Since "*requirements acquisition and product presentation often cannot be completely separated*" [BERGMANN2002], we need to enable some type of relevance feedback to further refine the query to match the searcher's information need. In order to quantify the searcher's interest in a given document, one must attempt to ascertain the searcher's focus of interest.

White ([WHITE2003]) accomplishes this by separating a document into various stages of representation, in which each stage increases the level of detail. The searcher is then enabled to drill-down, through the path of representations and acquire more information about the document's content. One could, for instance, create a consecutive path of representations for a document, by using Ingwersen's "*modes of presentation of objects*" [INGWERSEN2000]:

- **Bibliographic Data**: The most abstract representation offers easily accessible information (author, title, keywords) to facilitate recognition of relevant topics and concepts.
- **Summary**: The summary offers a slightly more elaborate synopsis of the document. Note that it is even possible to automatically generate an (indicative) summary from a given document [MINYENKAN2003].
- **Object**: Once the path of representations has been completely traversed, the searcher is presented with the document itself.

Obviously, more and different steps of document representation can be devised. The searcher is ultimately in control as to what representations he wishes to see and if he wishes to drill further down its path. When the searcher chooses to abort the path, it can be assumed that the information in the current representation was not personally relevant. The distance travelled down the path indicates the searcher's commitment to that document, and the personal relevance thereof.

When the searcher chooses to drill-down, the current representation can be assumed to be personally relevant. This mean that the terms in that representation can be used to extend or refine the query, allowing for a greater precision of the query's results. In this way, the separation of a document into various representations enables implicit relevance feedback, without interrupting the searcher during his search process. Based on this feedback, the set of documents offered to the searcher can be changed, to more accurately reflect his information need.

Intuitively, the most recently viewed path of representations can be assumed to be more relevant than previously viewed paths, since the searcher's concept drift (a drift in the nature of his interest) or due to the fact that the searcher may well be experimenting with the search-engines interaction.

7. Theory in Practice

"Multimedia scares me off." - Michael Nesmith

Information is used in a multitude of domains. Information can thus pertain many types of objects; be it a document (text), data about another object (meta-data), multimedia (i.e. a movie) or even an hypermedia (an interactive experience, i.e. interactive movie or video game). Thus far, the consolidated model has been based on 'pure' information-seeking within the scope of documents. The following research will however project the model on a *Universe of Discourse* (UoD), being the part of reality to be modelled. In this case, the UoD deals with today's entertainment industry, focussing on multimedia and hypermedia (in particular movies and video games).

A *movie* combines multiple media (motion pictures, sound) to communicate a message or story to the audience. The presentation of this message is limited only by the director's imagination, making some movies picture true works of art.

A Video game, like a movie, is also designed for the consumer's enjoyment. However, their design deviates from most movies in the sense that the consumer actively participates in the format, influencing its course of action. This makes video games non-linear, as oppose to the intrinsic linear movies. The addition of this responsive interactivity adds a new dimension to the sub-domain; different players might (more often than is the case with movies) have a different experience of the same product.

7.1. Information Needs

From observing the entertainment domain, several similarities in information needs can been identified. For instance, someone might be inclined to search for an object (either movie or video game) they experienced before, yet are unable to identify due to it only being a vague memory. Alternatively, a person found a given object quite satisfying and is looking for other objects that are *similarly* satisfying. Or, more specifically, someone enjoyed a rather satisfying object and is able to translate this experience to a set of intrinsic characteristics, which can then be used and tweaked to find objects that are more satisfying.

The above information needs reflect people with prior experience with the domain in question. However, someone can even ask people for advice in suggesting certain objects for themselves or others without prior domain knowledge. This can occur either with or without (directly) posing additional constraints in terms of external indices. An example for this would be parents asking what to buy their son for his birthday, while having no knowledge of the domain themselves. In this case, the searcher will want to provide enjoyable objects for their offspring, yet be advised about possibly offending content thereof.

The information needs throughout the entertainment domain can be expressed by discriminating the following typical information need constructions:

- Give me (a set of) characteristics of a given product
- Give me (a set of) products, *related to (i.e. similar to)* a given product.
- Give me (a set of) products, with the given (set of) characteristics

Note that the nature of the characteristics can be both internal as well as external.

Each of these categories are affected by several types of query components:

- Exact Match, referring to exactly match an explicit characteristic.
- Range Query, referring to a range of acceptable characteristics.

The search conditions can be compounded, to specify multiple criteria for multiple feature dimensions in a single information request.

When people look for mediated support, the expressiveness of their information need differs. This is mostly due to inconsistencies or gaps in their knowledge about the once experienced characteristics of the object (memory), or due to recalling only insufficiently discriminating characteristics (domain awareness) to help identify the object. So, each of the above information need constructions are subject to another parameter that affects the completeness of the search conditions. For example, a certain object's title is only vaguely recalled, in the form of certain keyword(s). Similarly, specifying an incomplete (in terms of discriminability) similarity matrix will not correspond with the searcher's expected feeling of similarity, thereby wrongfully suggesting that there are no similar objects.

Just to give a scent of some sample information requests solved by intermediation, here are some examples from a webforum ([FOK!]):

- "I'm looking for an old puzzle game. It somewhat resembles 'Supaplex' and 'Paganitzu', but I believe it featured monochrome display. All I can remember is that there was a time limit involved. It's not 'Rockford' or 'Boulderdash'."
- "I'm looking a game for the Playstation console. You're in a sport car, undertaking all kinds of missions. It also included blowing up some tanks with a laser.", later elaborated by "It unfortunately isn't 'Grudge Warriors' or 'Wild Metal Country'. The game I'm looking for is situated in a town instead, and somewhat resembles the game 'Driver'. You can't get out of the car."
- "I once had a karate-game on the Commodore 64 that had you walking around, fighting enemies. The amount of 'lives' you had was displayed at the bottom of the screen.", later elaborated with "No, it isn't Karateka: the game I'm looking for had bigger characters and the number of lives you had was displayed as a papyrus scroll as oppose to ordinary triangles".
- "I'm looking for a game, a Real-time Strategy Game. The publisher was from France, the style was Cartoon-like with a WWII style. I recall aircraft-carriers, tanks and bulldozers. You needed tankers to supply your troops with fuel."
- "I'm looking for a movie I saw in the 80's. I believe it dealt with aliens or robots on earth that looked like people. One of them was in love with an earth woman."
- "I'm looking for a movie, taking place during WWII. There's this American (I think played by Nick Nolte, but I'm not sure) who teaches a tribe of Indians oppressed by Japanese to fight back."

Though the above examples were taken from one forum, there are several forums dealing with this sort of intermediated discussion.

The information requests differ in both structure and characteristics. As is the case with the video games requests, some people actively remember the system that played the game, while others only remember a specific genre and visual feature to differentiate the relevant result from suggested ones. What is interesting to see is how these information requests evolve as other people respond to them. For instance, in the case of the karate-game, one person suggested a specific title, which was then deemed irrelevant by the searcher. This intermediated suggestion however even elicited more information from the searcher in terms of why the suggested game doesn't match his information need.

Also interesting to note is the fact that someone seems to remember a movie starring a specific actor in a specific plot. Despite the rather high information need awareness of this person, he/she seems to lack the system awareness to actually fulfil this information need.

Finally, it seems like this sort of intermediated searching is exclusively performed by human intermediaries. Additionally, due to its disorganised property, it is often performed on-line through unbounded (and even unstructured) channels like fulltext message boards, newsgroups or even live chat software. The downside of this is that whenever other people at a later time have a similar information need, they would either have to locate and sift through all previous discussion to find the answer, or present their information request anew, so people will be solving a similar problem all over again. Unfortunately, the latter option is considered best practice: examining unstructured data takes far too much time, especially when considering the sheer bulk of records available.

7.2. Problem Space Composition

In the field of entertainment, the most general problem space takes the form of the implicit information request, e.g. "I want to be entertained". In order to resolve this complex problem, one needs to decompose it into several smaller (manageable) problems. Following the divide-and-conquer principle, the smaller the individual problem, the less effort it takes to effectively solve it. The following illustration offers a schematic representation of a possible (albeit simplified) problem space construction:

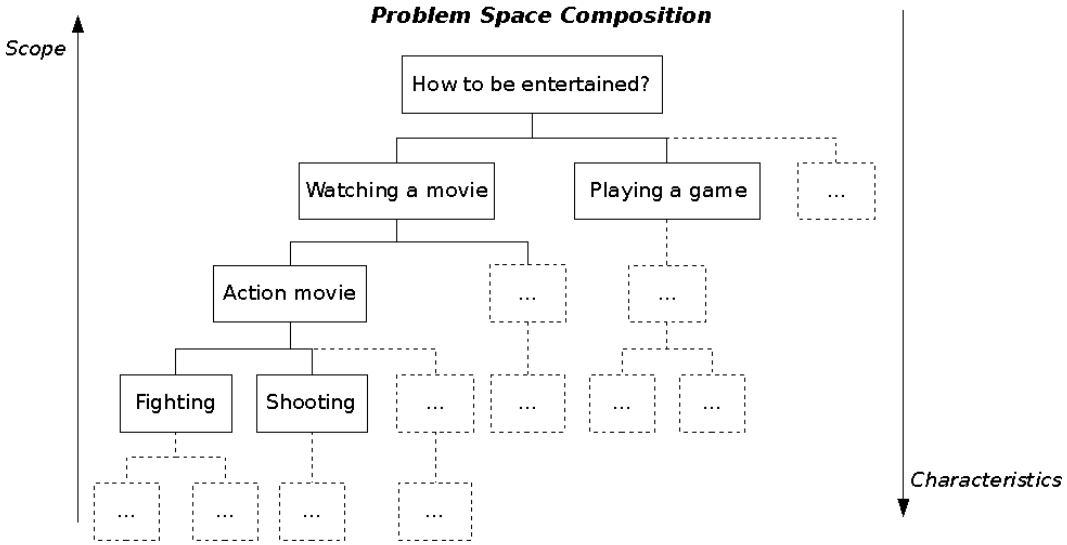

Illustration 12: Problem Space Composition

When drilling down the tree-like structure, it is noticed that the scope of the problem (the root node) gradually changes context and is translated to domain-specific characteristics. It is important to note that this observation is also related to the model as depicted in *Illustration 12: Problem Space Composition*; while drilling down, the problem space become smaller, yet require specific characteristics to be provided in order for them to be resolved. These characteristics can be of both internal and external nature, depending on the searcher's system awareness. Furthermore, the scope of the problem space corresponds with the searcher's information need awareness; the lower the awareness, the broader the problem space.

Because characteristics only come into play in problem spaces with a relatively small scope, searchers with a low information need awareness as well as a low system awareness (which corresponds with a *viscerally oblivious* searcher in *Illustration 4: Information Need Awareness vs. System Awareness*), are unable to drill down; they will require a higher information need awareness or a higher system awareness in order to do so.

Projected on the domain of video games, the Problem Space Context of a searcher trying to remember the game 'Tetris' might at some point resemble the following illustration:

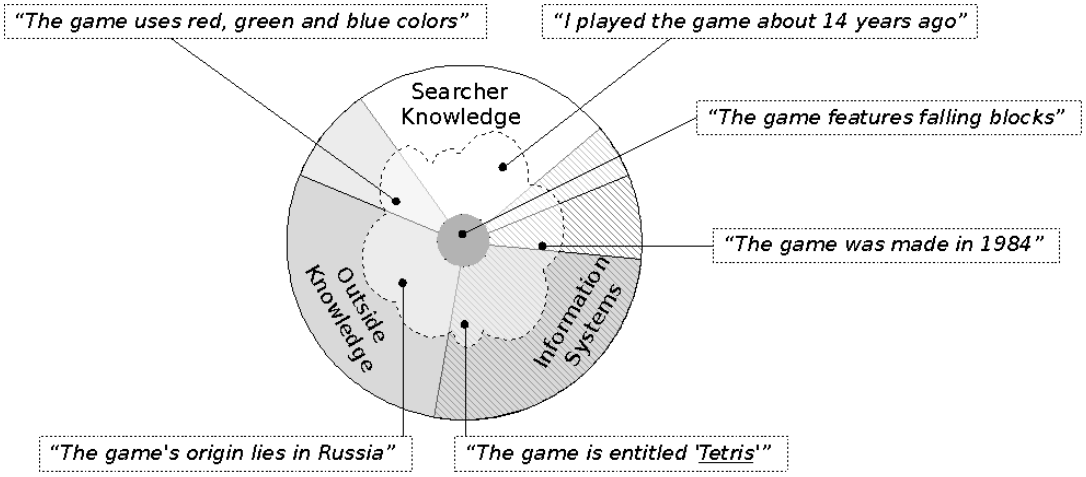

Illustration 13: Exemplary Problem Space

Note that this illustration differs from the Problem Space Context as it was introduced in *Illustration 1: Problem Space Context*. Firstly, the outside knowledge is a relatively big area, spanning about two-third of the complete circle. It has a relatively huge overlap with the information systems, indicating that a lot of the IS' information is known by the society of people familiar with the video games domain. Secondly, the unexplored area (originally denoted as φ) is not present in this situation, because the problem space deals with a rather well-known closed problem (the concept of which is created and thus intrinsically completely defined by man himself). Consequentially, the center of the above illustration is now formed by the intersection of searcher knowledge, outside knowledge and Information Systems instead. This area contains the general facts of the game, as shared by all partitions.

The Information Systems can be categorised as offering some obscure facts that are not really known by many (the exact year in which the game was made). Similarly, the searcher has unique knowledge in terms of him playing the game around 14 years ago. Although this fact might seem of little significance in its current form, it does allow for further processing: if the searcher has sufficient system awareness to know that an IS offers the ability to limit the results based on their time frame, it can be transformed into a restriction on this internal characteristic. So, the IS would only consider games that were made at least 14 years ago. Depending on the repository, this could be a key condition for discriminating among the games offered.

Typically, Information Systems deal with a model of a particular UoD. The model's properties depend on the knowledge and focus of the domain analyst responsible for the model; the analyst might model aspects of the UoD that are not relevant to the searcher. Similarly, some aspects relevant to the searcher might be overlooked by the domain analyst. An example of this is shown in the above illustration, where subjective facts about the experienced sensory (visual) perception of the object are only found as constructs in (searcher- and outside-) knowledge, and not in information. This phenomenon relates to the intrinsic property of man to make sense of the world around him. Using knowledge concepts known and experienced by him, the searcher tries to most accurately convey the wanted game's characteristics. In this case, the searcher clearly remembers the characteristic of the game using "red, green and blue colors". Being an external index, this aspect is not modelled in the Information System, because the domain model does not recognise it as one of its internal indices.

By employing our consolidated intermediation model to generate and store external indices for use in the Information Systems, several of the searcher-remembered experienced characteristics will also be present in the Information Systems-area, thereby allowing the searcher to specify this aspect in his request, to optimize the judgement of relevance. In this case the characteristic referred to is obviously the fact pertaining the "red, green and blue colors" being used; the fact that the searcher played it 14 years ago is not known by other people and will therefore never emerge from the meta-information, gathered from the outside knowledge through human computation.

7.3. Domain Object Identification

An object can be identified through either weak or strong identification. Strong identification involves the addition of an artificial key (standard name), which is unique for each object per design. Weak identification applies to identifying objects, based on each having a unique set of properties. The concept of identification can be incorporated in our model, stating that people who are searching for a specific object would be able to pinpoint this object through either strong or weak identification, corresponding to the compromised state. This compromised state implies a very high level of both information need awareness and system awareness (especially the case in strong identification) that may not be obtained by the searcher if he prematurely deems his information need satisfied prior to reaching the high level of system awareness.

The effectiveness of weak identification depends on the UoD: how many objects are present, which characteristics are established and how discriminative are these characteristics? For example, in both movie and video games domains, people are commonly inclined to refer to objects by their title as a way of identification. This assumes that a title of such an object would be unique, but objects with similar titles emerge as the UoD expands. For example, a movie titled 'Action' was released in 1980. This title was only unique until 1994, when another movie was released, bearing the same title. So although the concept of a 'title'-characteristic help eliminate most irrelevant results, it should not be treated as a method of strong identification. Instead, it should be used as a filter to narrow down the choices in a weak identification, augmented with a set of other characteristics to arrive at weak identification.

In practice, the expressiveness of a title to identify an object can lead to inconsistent information, being attributed to an incorrect object. For example, with websites offering visitors the opportunity to add additional characteristics (i.e. reviews/ratings etc.) to an entry, the visitor might be inclined to mistake one object for another, based on just the title as identifier.

In the light of the problem space context, the domain objects' characteristics are represented as the facts/truths of said domain. Such an object, due to weak identification, is stored as a set of characteristics (both internal and external) and placed in the problem space context. Note however that each

part in the problem space context (be it searcher knowledge, information systems or outside knowledge) might have an other way of identifying an object, depending on the depth (index-wise) and breadth (number of objects) of the repository of knowledge/information available.

For example, some person might describe and identify a certain game by remembering the characteristics of it being a "2-player board game", involving an "8x8 board", having "white pieces for one side, and black pieces for the other". For that person to identify the game through weak identification in this manner presupposes that said game is the only logical candidate for that description. That is, the searcher's knowledge is only aware of one game that fits the bill, let's say 'chess'. However, the game of 'checkers' also matches this description, although the searcher does not know this game.

Now, whenever the searcher comes across an IS recording both games, a conflict between the searcher's domain view and the IS' domain view will ensue. This conflict can be resolved by either party:

- The searcher acknowledges the fact that his own domain model is incomplete, and adjusts his model to that of the IS. Here, the searcher made a serendipitous discovery and learned that the domain is bigger and more complex than he anticipated. Thus, he acquired more system knowledge.
- The IS might extrapolate that the searcher originates from an environment where chess is played in favor of checkers. In this case, the IS judges to return the most likely (or popular) candidate in favor of addressing the information need, rather than providing more system knowledge.

7.4. Example ADP's

Combining *Illustration 12: Problem Space Composition* and *Illustration 5: Exemplary Awareness Development Paths* with the example domain will give a more concrete understanding of its applications.

When the start state of the ADP is *viscerally oblivious*, the searcher is unaware of the options available to his problem space of wanting to be entertained. The problem space scope is very broad, with the corresponding information need being too complex to describe. The complexity lies in having a low system awareness, along with a low information need awareness. Combining these values leads to a state of having too limited domain knowledge to effectively compose a request. This translates to a situation in which people can only describe their most abstract need, in this case something like "How can I be entertained?" For the next step, the following transitions are possible.

- Gaining information need awareness (following ADP1 – Taylor-esque): In order for an IS to address the information need, it must present the searcher with ways to drill down and decompose his complex problem space. This is done by establishing the most discriminative characteristics, and relaying these to the searcher. By doing so, the searcher passively gains system awareness. For instance, the IS might offer information about there being several options to entertain oneself, including watching a movie and playing a game. Based on the differentiating characteristics (i.e. outweighing passive entertainment against active entertainment), the searcher can assess the next level of his problem space, allowing him to drill down again.

- Gaining system awareness (following ADP2 – Experimental Search): The searcher will experiment with the IS, and obtain knowledge of the IS' interface and repository. In contrast to gaining information need awareness, this is an active process, initiated by the searcher himself rather than the IS. Reflected on the entertainment domain, the gained system awareness could also take the form of being able to distinguish passive and active entertainment, like is the case with gaining information need awareness. However, it also allows the serendipitous discovery of many other internal indices, which are even more discriminative in nature. This will ultimately allow the searcher to gain insight in the problem space composition, and drill down his problem space multiple levels at a time, as well as shifting his focus to neighbouring branches.
- Switching from gaining information need awareness in favor of system awareness (following ADP3 – Arrested Development): During this time, the searcher changes direction, wandering off the Taylor-esque path. During his search, he might have been confronted with a stimulus that tested his curiosity, causing his focus to drift. Though he gained a greater information need awareness, he now concentrates on obtaining more system awareness because he believes it will help him resolve the problem space faster.

The start state of the ADP does not have to be viscerally oblivious: it can also be *formalized oblivious*: the searcher has a pretty good understanding of his problem space and the intrinsic information need, yet possesses little to no system knowledge. Quite trivially, because the information need awareness has reached its peak, the only option for the searcher to evolve is to interact with the system to see how the IS can satisfy his information need. Because his knowledge of the IS' internal indices is low, the searcher will logically try and see if his known external indices are present in the system. Depending on the internal domain model, the external characteristics would either be a (partial) match to the internal ones, or provide insight in a possible relation to them. In the entertainment domain, this might pertain a domain expert switching to an unknown IS, which uses different characteristics than the one he's accustomed to. For example, some Information Systems in the video game domain only link a game to one broad genre, while others even distinguish additional sub-genres and themes; there's a difference in the characteristics' granularity.

Aside from formalized oblivious, the ADP can also start with high system awareness and low information need awareness, composing the so-called *viscerally cognizant* state. In this state, the searcher is an expert of the IS' interface and its repository, yet is rather visceral on his information need. This constitutes a situation in which the searcher can resolve any information need using the IS without a problem, yet fails to grasp the nature of his problem space and thus his information need. In this case, the searcher might somewhat be able to decompose his problem space, yet is unable to valuate the different options available to him: he's having a hard time weighing and comparing them in the context of his information need. This case is rather complex in terms of search behaviour: it presupposes a high system awareness, except having little insight in what the searcher actually wants.

8. Fulfilling the Information Needs

"Keep high aspirations, moderate expectations, and small needs" - William Howard Stein

Now that the problem space and search behaviour have become clearer, this chapter deals with fulfilling the information needs using both established Information Systems and a possible adaptation of our Consolidated Intermediation Model.

8.1. Established Information Systems

In the example domain, there's a variety of Information Systems. This chapter will highlight the most important ones, which are used as a reference point.

8.1.1. Movie Information System

In the domain of movies, several information systems exist that record various movies made throughout the years, the most elaborate of which is The Internet Movie Database [IMDB]. IMDB is an on-line database, dating back from 1990. Since then, it has undergone several changes in terms of repository and usability. Its interface allows for navigating a conceptual model, in which movies are generally composed of the following dimensions and related characteristics:

- Cast information (i.e. director(s), writer(s), actors, etc.), production information.
- Release information (i.e. release date, MPAA rating, alternate titles, etc.).
- Technical details (i.e. aspect ratio, sound mix, runtime).
- Plot information (plot summary, keywords, memorable quotes).

Navigating the IS can be performed by analytical search for title, names, companies, keywords, plots and movie characters. Contributing new movie entries, release information and technical details to the database is done exclusively by the IMDB staff. In terms of plot information, IMDB also distinguishes genres, yet they are a subset of (the most often used) plot keywords.

In cases where IMDB does not directly address the searcher's Information Request, the site's Message Boards can be used to get in touch with other visitors. IMDB's Message Boards are divided into several topical sections. As such, a board can cover a particular genre, locale, technology, or movie. The type of messages on these boards range from idle chit-chat to polls and questions in the hopes that other visitors can answer them. This latter case presents the visitor with a tool that allows their questions to be answered by other visitors of the site. These visitors would in turn each have a different background and knowledge with which to possibly answer the question, so one could state that someone responding on the Message Boards assumes the role of intermediary. More specifically, assuming multiple people respond to a single inquiry, the searcher will only have to specify his request *once* on the board, while retrieve results deemed relevant by a *multitude* of different people (intermediaries).

IMDB offers people to 'tag' movies with user-submitted keywords. The benefit of this approach is the fact that its descriptiveness is not limited to predefined values in the internal structure, making the database more dynamic in terms of supplying relevant information. However, the problem lies in the fact that a very large group of people has a very erratic train of thought due to conflicting concept definitions. The resulting tags are not formally defined and therefore prone to being partly ambiguous.

Furthermore, the tags are used inconsistently: people will mostly be inclined to apply newly introduced tags to newer movies, resulting in the older and more obscure movies having a weaker specification and thus a less elaborate index as opposed to the popular ones. Ironically, the fact that the latter movies are more popular already slightly indicates the fact that the majority of people already are aware of them, thereby eliminating the need for an extensive index. Perhaps as time goes by, the movies' popularity will diminish in favor of even newer movies. Only when that happens will the elaborate index have a purpose, yet at the same time constitute a bigger difference with the even older and poorly indexed movies.

8.1.2. Video Game Information System

In the domain of video games, several information systems exist that record video games, the most extensive and representative of which is MobyGames [MOBYGAMES]. Although the Internet Movie DataBase also records some video games, its video game database is far from complete, in comparison to MobyGames'. Furthermore, IMDB treats video games as specialized cases of movies, without actively separating them on the site. Therefore, people with an information need concerning video games would have to wade through their search results, separating the relevant games from the irrelevant movies manually. This relates to the difference in problem space scope, as detailed in *Illustration 12: Problem Space Composition*.

MobyGames' interface allows the searcher to specify the following dimensions in their search for relevant entries:

- Release information (i.e. release date, publisher, developer).
- Credits information (i.e. programmers, artists, etc), production information.
- Game Description (plot description, advertising blurb, game manuals).
- Genres and subgenres, themes.
- Technical details (i.e. platform, system requirements, supported features).
- Third-party Ratings (i.e. Kijkwijzer (Dutch), ESRB rating, etc).
- User-provided reviews of the game.

MobyGames differs from IMDB in the sense that it's more open for contributions: any user can contribute any kind of information, ranging from providing screenshots for the video games, to submitting a new video game to the database.

Similar to IMDB, MobyGames features message boards where visitors can talk about anything, and get help finding games they are looking for.

MobyGames does not offer keywords. Instead, they use a static mapping of genres and themes to index the video games. The unchanging nature of the index will help people familiarize with the IS, giving them a high system awareness. However, the downside is the fact that the index used might be too static to correctly differentiate between the many objects that exist in the domain. As time goes by, the nature of video games evolves. New genres and concepts are created, some of which fall outside the index MobyGames defined. To counter this, the values used for indexing should be revised and/or expanded. However, MobyGames chooses to use 'nearest matching' index terms to try and remotely relate the new concepts to the existing categorization, thereby stretching the definition of the categories used throughout the system. Additionally, this means that some genres will turn out to be trivially discriminative and should therefore be avoided. An example of this is the genre 'strategy': a multitude of games involves strategic thinking in some form, so offering this search dimension does not help to differentiate the games.

8.1.3. Current Information System Evaluation

The big downside of the current Information Systems is the fact that their domain models are quite static, and require people to think in terms of the underlying data structure. This is well suited for commonly used internal characteristics, yet recording the external characteristics that live in the users of the IS still poses a problem.

When evaluating the aforementioned Information Systems in the light of the information needs pertaining to the domain, the observation is twofold:

- The Information Systems can easily address problem spaces with a very low scope, provided the searcher has a clear understanding of the available (internal and/or external) characteristics and the role they play in his information need. In this case, having a common vocabulary ensures that information requests are unambiguous and match the inner structure, thereby facilitating the search. In other cases, if the searcher is able to express his information need in terms supported by the IS, the underlying internal structure will yield the relevant results (provided the IS delivers on soundness and completeness).
- The Information Systems can not easily address problem spaces of a higher scope. For example, request terms that do not directly relate to the IS' internal indices are often disregarded. Instead, searchers are forced to express their information need in terms the IS uses. Some Information Systems implemented counter-measures, such as free-form interfaces in the sense of contact forms, forums and message boards to help the visitor in any way they can. However, because of the lack of proper structure in these measures, it is hard for people to browse, especially when they lack both the time and the system awareness to know where to find the relevant information.

All in all, the reviewed Information Systems do not offer additional support for people throughout the stages of Exploratory Searching; they basically provide a graphical interface shell around an underlying data repository, and leave the rest to the searcher. The interface assumes a fairly high level of information need awareness and a moderate level of system awareness to be of any use to the searcher. If that fails, the searcher can get more help through more conventional channels like message boards.

8.2. The Consolidated Intermediation Model

Information Systems designed with the Consolidated Intermediation Model in mind will be able to exhibit all the characteristics present in the analysed Information Systems. After all, the model offers all the available internal characteristics present in those systems, yet augmented with external ones.

The addition of reliable external characteristics to be used as indices requires quite some extra data to process. However, every additional fact about a stored entity in the IS can possibly contribute to helping searchers find the things they want. Furthermore, because the IS' design has explicit and structural support for the addition of external indices, the resulting IS will actively involve them in the interaction with the searcher, as oppose to forwarding the searcher to an unstructured stack of data to wade through himself.

Finally, by incorporating an intermediation module, the Consolidated Intermediated Model-based IS should be capable of detecting the searcher's place in his Awareness Development Path and offering support, custom-tailored to the searcher's need.

9. Conclusion

"Reasoning draws a conclusion, but does not make the conclusion certain, unless the mind discovers it by the path of experience." - Roger Bacon

By correlating the various concepts surrounding the searcher's problem space (e.g. searcher knowledge, Information Systems, outside knowledge), we were able to create a model in which the exact nature of their underlying relationships became apparent. Furthermore, the overlapping regions resulted in a comprehensive delineation of the various categories of information and knowledge, with respect to the searcher's knowledge. Using this model, we were able to more accurately express existing concepts (e.g. personal and semantic relevance) and the differences between them.

Taylor's definition of the various information need levels was not completely outlined and thus subject to interpretation. By expressing these need levels in terms of Belkin's information seeking dimensions, this definition was made more concrete. In doing so, we were able to connect various models, each describing a set of search strategies, to Taylor's information need levels. First of all, this enabled us to see how a certain state of information need awareness resulted in the corresponding search strategy, as employed during the search process. Second of all, this allowed us to see that each of the models described a similar set of search strategies, using different terminologies. Using this insight, a unified set of search strategies was proposed.

The addition of the system awareness dimension (aside from the information need awareness), expanded the set of dimensions in which Taylor's need levels could be described. Using this additional dimension, we were able to differentiate the compromised need level from the preceding need levels (i.e. visceral, conscious, formalized). This resulted us to conclude that the compromised level was in fact not a separate need level, but rather an expansion upon the formalized need level with additional system awareness. Furthermore, the system awareness dimension resulted in a Awareness Development Path (ADP) that describes searchers' progression between the two forms of awareness.

The unification of existing (overlapping) studies into the actions and operations performed by searchers resulted in a single search flow chart. This schematic representation of the possible sequence of search operations describes the many ways in which people are able to acquire, consume and filter information.

By acquiring inspiration from the practical (human-to-human) intermediation models and by translating our observations derived from those practical models, we have composed a consolidated intermediation model to enable digital IR intermediation. The introduction of human computation in order to acquire human perception information allows an IS (built according to the consolidated intermediation model) to interact with searchers and perform requirements acquisition in terms derived from the human perception on documents. This addition of external indices to supplement the internal indices allows for a more true to life interpretation of the searcher's information need. Combined with the relationship-map, which allows an IS to propose related additional keywords much like a human intermediary would, we believe the consolidated intermediation model to be more proficient (than existing search engines) at guiding searchers during exploratory search.

Although, in order for the external indices derived from human computation to be most effective at assisting searchers in conveying their information need, one first needs to obtain a sufficient degree of information (in the form of external indices) through human computation. Depending on the size of the repository and the willingness of people to participate in this type of implicit data-acquisition, this process will take some time. However, we feel that the added value that the external indices offers this type of intermediation, ultimately outweighs the (initial) effort it requires.

10. Recommendations

"Why give advice?... A wise man won't need it.
A fool won't heed it." - Origin unknown

This chapter offers several recommendations for future research to expand the research presented in this document.

- The findings of this research would seem fit to be adopted by projects dealing with *Web 2.0*, referring to on-line collaborative platforms. The success of initiatives like Wikipedia [WIKIPEDIA] show that there is a momentum towards these community-based collaboration. Therefore, the thought of incorporating Human Computation in such a project would be the next logical step.

- The Consolidated Intermediation Model can be used as a foundation for all sorts of implementations in a variety of domains. The deployment can range from simple single-domain Information Systems to complex Information Systems, spanning multiple domains. By cleverly making use of the hierarchy of the relationship-map, multiple applications incorporating the Intermediation Model would be able to share their Human Perception Information datastores, as well as their search functionality. This could ultimately unlock a multi-tier architecture to deal with compound multi-domain problem spaces.

- Seeing that there's already quite some user-submitted information available on web-forums and newsgroups, perhaps it would be possible to construct some sort of spider application, capable of filtering out the relevant information and disposing of the enormous information overhead that exists on the web today. The relevant results would then be considered valuable sources for external characteristics.

Bibliography

AGMRC2007: "*Relationship Selling: The Path to Sales Success*". 2007. http://*www.agmrc.org/agmrc/business/operatingbusiness*
BAEZA-YATES1999: Ricardo Baeza-Yates, Berthier Ribeiro-Neto, "*Modern Information Retrieval*". , 1999.
BELKIN1993: Nicholas Belkin, "*Interaction with texts: information retrieval as information-seeking...*". Rutgers University, 1993.
BELKIN1994: Nicholas Belkin, "*Cases, Scripts, and Information-Seeking Strategies*". Rutgers University, 1994.
BELKIN2005: N. J. Belkin, "*Anomalous State of Knowledge*". , 2005.
BERGMANN2002: Ralph Bergmann, "*Acquiring Customers' Requirements in Electronic Commerce*". University of Hildesheim, 2002.
BRUCE2005: Harry Bruce, "*Personal, anticipated information need*". University of Washington, 2005.
FOK!: "*Fok!*". 2007. http://*http://forum.fok.nl*
IMDB: "*The Internet Movie Database*". 2007. http://*http://www.imdb.com*
INGWERSEN2000: "*Users in context*". 2000. http://*ei.cs.vt.edu/~cs5604/f01/lectures/ingwersen.pdf*
KUHLTHAU1993: Carol Kuhlthau, "*Seeking meaning*". Rutgers University, 1993.
LARSEN1997: Jan-Helge Larsen, "*P-R-A-C-T-I-C-A-L, a step-by-step consultation model*". Department of General Practice, 1997.
MARCHIONINI1995: Marchionini, "*Information seeking in electronic environments*". Cambridge University, 1995.
MINYENKAN2003: Min-Yen Kan, "*Automatic Text Summarization As Applied To Information Retrieval*". Columbia University, 2003.
MOBYGAMES: "*MobyGames*". 2007. http://*http://www.mobygames.com*
NIEDZWIEDZKA2003: Barbara Niedźwiedzka, "*A proposed general model of information behaviour*". Institute of Public Health, 2003.
SHIMAZU2002: Hideo Shimazu, "*ExpertClerk: A Conversational Case-Based Reasoning Tool...*". Internet Systems Research Laboratories, 2002.
SONNENWALD2001: Diane Sonnenwald, "*Evolving perspectives of human information behavior: contexts, situation...*". University of North Carolina, 2001.
STOJANOVIC2005: Nenad Stojanovic, "*On the role of a user's knowledge gap in an information retrieval process*". University of Karlsruhe, 2005.
TAYLOR1968: Robert Taylor, "*Question-negotiation and information seeking in libraries*". Lehigh University, 1968.
VANVLIET2007: Mario van Vliet, Theo van der Weide, "*The sales dialog as a model for information retrieval*". Radboud University Nijmegen, 2007.
VONAHN2004: Luis von Ahn, "*Labeling images with a computer game*". Carnegie Mellon University, 2004.
VONAHN2006: Luis von Ahn, "*Games with a purpose*". Carnegie Mellon University, 2006.
WEBLIMINAL2006: "*Directories and Virtual Libraries*". 2006. http://*www.webliminal.com/search/search-web04.html*
WEICHOO2000: Chun Wei Choo, et al, "*Information seeking on the web: a model of browsing and searching*". First Monday, 2000.
WHITE2003: Ryen White, Joemon Jose, Ian Ruthven, "*Adapting To Evolving Needs: Evaluating A Behaviour-Based Search Interface*". University of Glasgow, 2003.
WHITE2005: Ryen White, et al., "*Exploratory search interfaces: categorization, clustering and beyond*". University of Maryland, 2005.
WIKIPEDIA: "*Wikipedia*". 2001. http://*http://www.wikipedia.org/*
WILSON1997: Tom Wilson, "*Information Behaviour: An interdisciplanary Perspective*". University of Sheffield, 1997.
WORDNET2007: "*WordNet, a lexical database for the English language*". 2007. http://*wordnet.princeton.edu/*