Project Plan

Colofon

Document: Project Plan
Version: 0.26
Date: 06-05-2007

Student 1
Name: Daniël Rutten
Studentnr: 0424625
Afstudeernr: 562
E-mail: DanielRutten@student.ru.nl

Student 2
Name: Jeroen Bakker
Studentnr: 0421898
Afstudeernr: 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. M. van Vliet
Address: Toernooiveld 1, 6525 ED Nijmegen

1. Introduction

In today's society, there is an abundance of information available to anyone who is interested in it. This virtually unlimited availability of large quantities of information about any number of topics introduces a serious dilemma: how does the consumer manage to locate the information that he is after? A similar question can be posed from the point of information publishers who want their published product to be found by interested parties: how can we ensure that the consumer finds the information most relevant to him, from the set of products that we have made available?

Information can pertain many types of objects; be it a book (text), data about another object (meta-data), a movie (multimedia) or even a video game (hypermedia). If the object discussed is a valuable resource for a number of consumers, it qualifies as a *product;* each product has a unique set of characteristics by which it can be identified. However, consumers generally do not always succeed in choosing the right set of characteristics to specify the product(s) they are looking for. Their ability of formulating a search query depends on the information need level [TAYLOR1968] they are subjected to.

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 intrinsic need level. 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*, in which the searcher is initially unable to formulate a precise query, and is merely able to recognize characteristics of the wanted resource. Any support given to this type of search process will need to encompass the entire length of the searcher's journey, going from the so-called 'visceral awareness'-level to the 'compromised awareness'-level. The main research question that can be derived from the aforementioned problem is as follows:

> *"How can a searcher best be assisted in refining his information need level, from the visceral to the compromised level, aimed towards guiding the searcher to the most relevant resource(s) available from a given domain?"*

Because of the state of current research being done in this direction, this question is best answered in the shape of a new Information Retrieval model, which should be applicable to a wide range of domains. This new model should guide and assist searchers on their path to the desired resource(s), gradually refining the searcher's query by encompassing the consumer's various need levels.

2. Planning & Organization

This study will be conducted in-house, by Jeroen Bakker and Daniël Rutten; both students of the Radboud University in Nijmegen, the Netherlands. The students' progress is supervised by prof. dr. ir. T. van der Weide and prof. dr. M. van Vliet, both professors at the Radboud University. To facilitate supervision and communication, the aim is to initially schedule a weekly communication, during which recent topics and progress is discussed. In practice, the way in which this communication is performed, be it through meetings or email, will be determined on a week-by-week basis. The duration of the study is divided in the following phases:

- **Orientation**
  This phase focuses on formulating a research question by (broadly) exploring the IR process (i.e. Process models, IR techniques, some cognitive aspects), to allow for in-depth research in the next phase. This phase results in the Project Plan, which conveys the purpose of the study, and gives an overview of the study in terms of planning and organization.
  Deliverables: Project plan, Repository of existing models/techniques

- **Research**
  The Research phase deals with the in-depth study of various IR models and techniques. Among these, we will study the Sales Dialog model [VANVLIET2007]. From these studies, a better understanding of the domain emerges, allowing for the composition of a new model for Information Retrieval (possibly expanding or complementing an existing model). This will be the main focus of the thesis. As an extension to the research, each student will conduct an individual area of research, relating to the new model. The two individual research topics will include the validation of the model, both theoretically and experimentally, respectively. The latter should also allow for a more profound testing of the theory's validity, since this would project it onto a real-world scenario.
  Deliverables: New theoretical model, Results of theoretical and experimental validation

- **Thesis**
  Finally, the thesis will be written. Because this study is conducted by 2 students, there is a requirement for 2 separate theses. However, because both cover the same research trajectory, the overlapping study will result in a common section as well. Ultimately, each thesis will contain a common section (dealing with documenting the shared aspects) and an individual section (aimed at documenting aspects pertaining the individual students' assignment).
  Deliverables: Thesis (2x)

- **Finalization and Presentation**
  This phase deals with preparing the thesis presentation. Additionally, it offers a buffer to help counter any delays in the execution of the study.
  Deliverables: Presentation

The global planning (with Academic Holidays marked in gray):

		Wk	*Date*	*Mon*	*Tue*	*Wed*	*Thu*	*Fri*
Orientation	**Project Plan**	**1**	12 Feb - 16 Feb					
		2	19 Feb - 23 Feb	Carnival				
	Breadth	**3**	26 Feb - 2 Mar		Meeting			
		4	5 Mar - 9 Mar		Meeting			
	Study	**5**	12 Mar - 16 Mar		Meeting			
		6	19 Mar - 23 Mar		Meeting			
		7	26 Mar - 30 Mar		Meeting			
Research	**New model**	**8**	2 April - 6 April		Meeting			Good Friday
		9	9 April - 13 April	Easter Monday	Meeting			
		10	16 April - 20 April		Meeting			
		11	23 April - 27 April		Meeting			
		12	30 April - 4 May	May Holiday				
	Individual	**13**	7 May - 11 May		Meeting		Dies	
		14	14 May - 18 May		Meeting		Ascension	After Ascension
		15	21 May - 25 May		Meeting			
		16	28 May - 1 Jun	Pentecost	Meeting			
Thesis	**Common**	**17**	4 Jun – 8 Jun		Meeting			
		18	11 Jun - 15 Jun		Meeting			
	Individual	**19**	18 Jun - 22 Jun		Meeting			
		20	25 Jun - 29 Jun		Meeting			
		21	2 Jul - 6 Jul		Meeting			
Finalization	**Presentation**	**22**	9 Jul - 13 Jul	Summer Holiday				
		23	16 Jul - 20 Jul	Summer Holiday				

Bibliography

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.
STOJANOVIC2005: Nenad Stojanovic, "*On the role of a user's knowledge gap in an information retrieval process*". University of Karlsruhe, 2005.
INGWERSEN2002: Peter Ingwersen, "*Information retrieval interaction*". Royal School of Library and Information Science, 2002.