Implicit and Explicit User Modelling Techniques for Interactive Visual Knowledge Exploration

by

Thomas Johannes Hengster

Dissertation

Presented to the

University of Dublin, Trinity College

in fulfillment

of the requirements

for the Degree of

Master of Science in Computer Science

University of Dublin, Trinity College

September 2009

Declaration

I, the undersigned, declare that this work has not previously been submitted as an exercise for a degree at this, or any other University, and that unless otherwise stated, is my own work.

_____

Thomas Johannes Hengster

3rd September 2009

Permission to lend and/or copy

I, the undersigned, agree that Trinity College Library may lend or copy this thesis upon request.

_____

Thomas Johannes Hengster

3rd September 2009

Acknowledgments

I would like to thank my supervisor Dr. Owen Conlan for his patience, invaluable assistance and guidance during the course of this project.

Further I would like to thank Cormac Hampson for his constant support, advice and time.

I also would like to thank my family for all the encouragement throughout the last year.

Finally, I would like to thank everyone from the Networks and Distributed Systems course. Keep up the good work!

Thomas Johannes Hengster

University of Dublin, Trinity College

September 2009

Abstract

This work investigates the challenges when modelling a user's intent as he interactively explores a knowledge space and correlating it with subjective expertise about that space. As this exploration is visual in nature the user is not burdened with unnecessary modelling and information seeking concerns in order to maintain flow and immersion.

A balance between explicit and implicit modelling techniques needs to be maintained to ensure the user has a contiguous experience. Furthermore it offers an alternative to collaborative filtering. It does not rely on peer ratings, but independently draws conclusions from continuous user interactions with the system.

As a result, it does not suffer from missing or incorrect pre compiled facts and correlations. Nor does it suffer from the sparsity problem that many collaborative filtering systems exhibit. The personalised nature of the interactions and the user's subsequent discoveries presents a need for detailed, yet potentially quickly changing user models. This also requires carefully balancing and throttling the influence of related semantic attributes.

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Overview

This chapter introduces the research question, its motivation and how it was approached. More specifically it will show how user modelling, implicit and explicit personalisation techniques combined with expertise and a compelling user interface can help exploring immense sets of data [**1**].

1.2 Motivation

The internet is a tremendously useful source of information. However, the universal problem encountered is: Creation, conversion and collation of huge volumes of data resulted in a vast amount of accessible information. It can be difficult to accessed and explore this large volume of information. With the appropriate knowledge and effort a user may retrieve what he is looking for. However, users usually have to pay a high price in the form of spending a lot of time and effort. This applies particularly to media related content [**2**], where the non-textual and often subjective nature makes it difficult to analyse and categorise automatically.

Films are a typical example of such non-textual media. In the film domain there are many different websites and services that try to improve access to structured information. These sources comprise detailed facts about films including cast, retail information as well as connections to other films. Especially the last part is significant. Imagine you want to watch a film: The first step in the process would be to pick one. This may sound rather trivial, but in most cases there are many different competing factors when picking a film. For example, how well it rates it critics may be a factor for one user, but the popularity of the directors of may be more important for another. If you don't have a specific film in mind, it is very likely you have at least a rough idea about what you prefer and what not. But where and how should you start your search? As we will see below, personalised use of expertise can help to overcome these issues.

There are quite a few systems like IMDb [**3**], MovieLens [**4**] or Rotten Tomatoes [**5**] that offer assistance. Many use precompiled recommendations of films related by

genre, similar actors or topic. Along with user-generated content like ratings or reviews these help to improve the burden of finding a suitable film. Most of them rely on fairly similar types of algorithms to extract and compute recommendations.

Due to the size of the underlying information even domain experts may encounter limits when maintaining single or distributed sets that tend to be barely comprehensible [**6**]. A manual solution is most often not feasible. But using different types of expertise encoded in abstract rules might help. It is crucial to have better techniques to enable superior understanding of domain data. Thus it is essential to facilitate concepts of information querying when trying to access and extract knowledge.

KDDM (Knowledge Discovery and Data Mining) techniques give the opportunity to reveal relationships and trends within large collections of data. This applies even if these attributes have not been explicitly encoded before [**7**]. Correlations usually rely on numerical, neural network, case-based or methods based on probability [**8**]. These types of approaches are rather complex. Users are often not willing to spend time and precious resources on formal methods of knowledge seeking [**9**].

In addition to that, many times a user's impetus, goal or expertise is ignored by these systems. As the human role is underemphasized the sole focus is set on the data and its structure [**10**]. Even if a user has a comprehensive understanding in a specific domain it might be difficult to find what he is looking for. The amount of content is too huge. This results in some sort of cognitive overload and has to be prevented. It can be achieved by reducing the quantity of information that impacts on a user, for example via splitting results and tasks iteratively. It is important to only present what is most valuable with regard to the circumstances [**6**].

This dissertation looks at how modelling a user's intent can aid user interactions with a system in order to personalise exploration of media. Human-computer interaction in combination with user modelling is the key. A tightly bound query and response paradigm enables intelligent interaction and allows managing the gap between a user's intent and a system's response [**11**]. A user often does not know in the beginning what he's looking for, i.e. he lacks having the right keywords, or does not really know what he wants, e.g. "Show me films I might like!" [**12**].

With the help of domain experts we bring in different subjective perspectives in form of semantic attributes which have been developed as part of the Semantic Attribute Reconciliation Architecture [**13**] in order to enrich the underlying meta information [**14**]. Implicit weightings and explicit interactions help modelling a user and balance what he is presented with while exploring a specific domain.

1.3 Research Statement

This research investigates whether implicit and explicit user modelling techniques can be appropriately applied to balance different subjective expert perspectives to aid exploration of media resources. It evaluates how this approach improves the exploration of large volumes of information where sufficient metadata is obtainable.

1.3.1 Objectives and Goals

In order to achieve this research statement the following specific objectives have been defined:

1. Analyse and understand State of the Art:
   1.1. Various KDDM (Knowledge Discovery and Data Mining) and current developments of information retrieval techniques.
   1.2. Differences between implicit and explicit user modelling.
   1.3. Different film recommendation systems.
   1.4. Different visual exploration systems and interaction metaphors.
2. Retrieve and organize domain data from various dissimilar sources and aggregate these rationally.
3. Investigate use of visual paradigms to convey information and the relationships between different pieces of information.
4. Survey challenges of modelling a user's intent as he interactively explores a knowledge space. This involves:
   4.1. Manage a user model that not only captures a user’s preferences, prior knowledge and goals, but also attributes some sort of certitude to these values that will change based on observed behaviour.
   4.2. Reconcile a user’s interactions with constant updates to his model.

5. Design and implement a compelling, yet easy to use user interface to entertain and further motivate exploration. This is key in the process of modelling a user's intent.
6. Evaluate the prototype, its user interface and implementation which is closely related to a prototype architecture called SARA (Semantic Attribute Reconciliation Architecture) [**15**] [**13**]. Furthermore assess exploration results and related user modelling based on silent and open feedback from expert users in addition to generic user trials.

1.4 Approach

When you are asked to find any new media, for example films, the first question that comes up is where to start. Especially when you should be looking for films you have never heard of. Implicit bounds and connections are not necessarily apparent to the information seeker. In a continuous progress user interactions define agreement or disagreement with the presented data. This helps the system to adjust the user model and refine its suggestions. The implementation can be tweaked to tolerate rapid focus changes while also harnessing a "short tail". The client prototype closely interacts with SARA [**13**]. The Semantic Attribute Reconciliation Architecture as well as its underlying process model support users exploring various data sources for any kind of useful information. It allows users to employ expert knowledge and utilises semantic attributes that make consolidated queries over several sources possible. It is important to mention that in order to benefit from this approach you need to have enough variation from disparate information sources. Another step in the process is to define interesting semantic attributes which are supported both in the client and SARA and therefore assist in bridging the semantic gap. The System learns from user interactions and presents new information he wasn't even aware of or didn't really know that he was interested in.

The chosen domain for this project is films. This allows for users being quite familiar with it to further support evaluation. It also simplifies initial data retrieval and visual appeal. Yet, with slight adjustments regarding to the user interface, any other domain could be incorporated as well. This would be particularly helpful in areas with little clique-based filtering information.

1.5 Proposed Contribution

This work will investigate the challenges when modelling a user's intent as he interactively explores a knowledge space and correlating it with subjective expertise about that space. As this exploration is visual in nature the user is not burdened with unnecessary modelling and information seeking concerns in order to maintain flow and immersion. A balance between explicit and implicit modelling techniques needs to be maintained to ensure the user has a contiguous experience. Furthermore it offers an alternative to collaborative filtering [**16**]. It does not rely on peer ratings, but independently draws conclusions from continuous user interactions with the system. As a result, it does not suffer from missing or incorrect precompiled facts and correlations. Nor does it suffer from the *sparsity* problem that many collaborative filtering systems exhibit [**17**]. The personalised nature of the interactions and the user's subsequent discoveries presents a need for detailed, yet potentially quickly changing user models. This also requires carefully balancing and throttling the influence of related semantic attributes.

1.6 Thesis Structure

After this introductory chapter, which provides an overview and brief details, the work continues with Chapter 2 where the State of the Art relevant to this dissertation's topic is reviewed. It gives closer looks at implicit and explicit user modelling techniques as well as KDDM, Flexible Querying and Dataspaces. Another important sub section informs about the prototype framework called SARA (Semantic Attribute Reconciliation Architecture) along with its process model. SARA had been designed to allow expert users to employ knowledge in the form of semantic attributes. Moreover, chapter 2 details about various existing (film) recommendation systems and visual exploration techniques. It points out primary features and criteria of how the systems function referring to ranking, filtering or visualisation paradigms.

Chapter 3 describes the Design of the system. It reassesses State of the Art and then introduces and explains the *Film Domain Exploration Client*. As a part of this, it promotes the application's behaviour and mechanisms. It defines the development process, main interaction metaphors, the exploration paradigms and how these are

related. It also explains the nature of semantic attributes, emphasizes on their importance and makes clear why they give a deeper meaning.

Chapter 4 deals with the software implementation. The project's implementation is examined and comprises sections about what directed the design and how the client was realised. On top of that, problems that have been encountered are pointed out.

In Chapter 5 the evaluation of the application is conducted. The results of the user experience and modelling are assessed. It will show whether participants have seen a benefit, and how and which design decisions or semantic attributes caused what effect. Improvements with regard to the prototype and the approach in general are suggested in addition.

The dissertation winds up with Chapter 6 where a conclusion is drawn. More specific information about the evaluation success and reviews which objectives have been achieved and which haven't. The client's applicability for other domains is briefly discussed. A section about possible future work finalises this chapter which is then followed by bibliography, abbreviations and appendices.

2 State of the Art

2.1 Overview

This chapter comprises the State of the Art that is significant to this dissertation. First of all I lay out details about implicit and explicit user modelling. This is to get an idea about the underlying research in KDDM, Flexible Querying and Dataspaces. Next is user modelling and personalisation. Then I discuss SARA – the Semantic Attribute Reconciliation Architecture – which plays a major part in the approach taken. Its implementation and how it facilitates employing expert knowledge within the system architecture is also a part. After that, a close look at current existing (film) recommendation systems, especially at their main features with relation to this work is conducted. This allows seeing what is available and where potential strengths and weaknesses are present. Then, an explicit review in the area of promising visualisation techniques follows. The chapter rounds-up with an analysis and conclusion.

2.2 KDDM, Flexible Querying and Dataspaces

"... Knowledge Discovery is the most desirable end-product of computing. Finding new phenomena or enhancing our knowledge about them has a greater long-range value than optimizing production processes or inventories, and is second only to task that preserve our world and our environment. It is not surprising that it is also one of the most difficult computing challenges to do well..." [**18**].

Due to the subject of this dissertation, both areas of Knowledge Discovery and Data Mining as well as Dataspaces are inherently important. Dataspaces are containers for domain specific data and thus loosely related collections of heterogeneous information. They can be found in enterprises, at home or in libraries. Organising and managing these different data sources through efficient methods is the aim of Dataspace Support Systems (DSSPs). They do not require unifying semantic heterogeneity and any of these services can be supplied instantly on multiple sources. As a result, it does perfectly co-exist with data integration systems and presents base functionality over information sources that are ignorant about their level of integration [**19**].

KDDM is a very dynamic research and development area that requires stable foundations. Current technological progress permits storage and access of huge amounts of data at almost no cost. In first place, the main benefit lies not in just having the data but being able to process it, finding interesting correlations by using statistical analysis and inference [**20**]. A survey about major process models by Kurgan and Musilek [**7**] describes those in use by the KDDM community. Any KDDM process model consists of a set of sequential steps including loops and iterations that have to be followed. A model contains procedures that have to be performed in each step: To plan, execute and reduce cost. Not surprisingly this results in a variety of different KDDM models both in academia and industry. Examples are the models proposed by Fayyad et al. [**21**], Cabena et al. [**22**], Anand & Buchner [**23**] [**24**], CRISP-DM [**25**] and Cios et al. [**26**], just to name a few of the most renowned ones. The key difference amongst them lies in the proposed number and range of their explicit steps. From this the authors extracted a six step generic model which consolidates the information merged from the others.

It is significant that previous research found that KDDM is extensively used but most often very complex to engage with. It frequently results in data dredging, discovering of too much insignificant knowledge while on the other side, user concerns are ignored [**7**] [**9**]. To accommodate for this, semantic relationships are becoming more and more central - especially between dissimilar resources. Automatic concluding of a range of inherent semantic features to support data mining can be found in a prototype built at Accenture Technology Labs. It allows for less human interaction and effort [**27**]. Little et al. [**28**] found that the process of discovering relationships within metadata essentially relies on its quality and compatibility.

In the field of flexible querying, crucial developments like Ontogator and OntoViews have emerged. They are concentrating on ontology based approaches to supply more competent search results. A lack of sophistication and a steeper learning curve for the users however, prevents widespread applications [**29**] [**30**].

Next, I will detail about a framework that is closely connected to the issues outlined above. Most importantly it offers promising ways to deal with the downsides.

2.3 SARA

2.3.1 Process Model

The Semantic Attribute Reconciliation Architecture (SARA) as well as its underlying process model as in User-Centric Exploration of Heterogeneous Information Sources [**15**] aims to support users exploring various data sources for useful information. It further allows drawing interesting relations in represented data that has not explicitly been encoded. This behaviour bases on the underlying processes that enable users to employ expert knowledge. This expert knowledge utilises semantic attributes that help consolidated queries over several sources at the same time. This type of personalisation assists bridging the semantic gap. That way, low level metadata can be controlled through semantics meaningful to users. SARA actually resides as a semantic intermediary between the end user or a client application and the core information sources.

The necessary processes within have been adapted from the generic six step KDDM process model specified by Kurgan and Musilek [**7**]. In SARA, data mining - the fourth step of the model – has been substituted with personalised querying. This makes it possible to introduce personalisation and encoding of expert knowledge to the process and subsequently SARA unifies and benefits from both paradigms [**15**].

The six steps of the SARA Process Model as shown in Figure 1 are as follows:

1. **Application Domain Understanding:**

   Learning the goals of the end-user and defining them in terms of the domain.

2. **Data Understanding:**

   Locating useful subsets of information sources and identifying any issues regarding quality or accessibility.

3. **Data Preparation:**

   Preprocessing of information that it meets the requirements of SARA.

4. **Personalised Querying:**

   Automated querying of underlying data sources according to the user's context.

5. **Evaluation**

   Results that are returned are visualised in a way to help users interpret the information and re-run queries.

6. **Knowledge Consolidation and Deployment:**

   Presenting generated knowledge in a human centred way.

Just like within the generic KDDM models each step builds on top of the other in a linear way. Feedback loops to any previous steps are possible and encouraged if a better outcome is prospective.

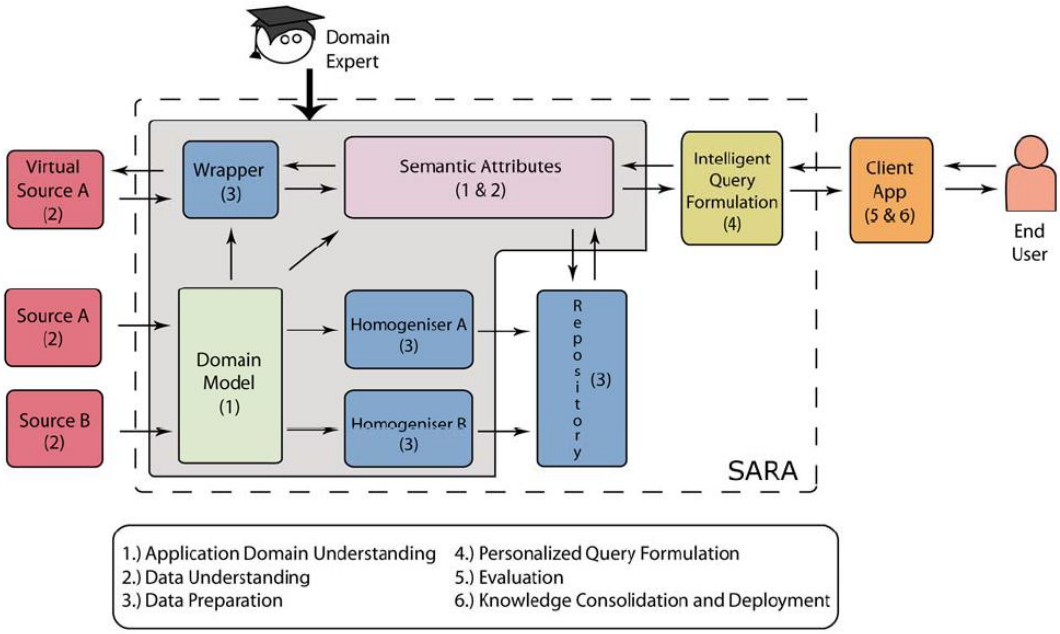


Figure 1: Generic KDDM process model as adapted and applied to SARA [**31**]

The domain expert has to build a model that includes key concepts of the domain that contains a homogeneous vocabulary set and data structure. The incorporated metadata from the external sources are conformed to these. This results in a fairly static but comprehensive structure that does not expect frequent changes.

Homogenised sources are transformed to fit the standardised domain model, but need not inevitably be merged or related to each other. This is the job of semantic attributes. They take care of combining and relating disparate sources. As such, they – and not the domain model – characterise potentially interesting trends in the data.

In the data understanding stage useful and preferably heterogeneous information sources have to be acquired and selected. These sources can be static or dynamic. However, they have to be structured or at least semi-structured so that their schema and semantics are to be understood by SARA. Flaws regarding quality, missing values and accessibility and their impact on the user's goals must be verified.

The next step, data understanding, deals with the harmonisation of input sources. This includes looking at semantic matching as well as syntactic and structural normalisation. An API makes it possible that external applications can interface with SARA and benefit from its functionality. The communication process exhibits that SARA presents what semantic attributes it provides. This supports users when defining substantive queries in relation to them. After this personalisation has taken place, the semantic attributes are reconciled into semantically aware search queries. Expert knowledge is encoded using semantic attributes. This brings extra semantic richness and more new ways of measuring things than just relying on statistical data that is not understood by the system.

2.3.2 Semantic Attributes

Semantic attributes represent discrete characteristics of a domain encoded by experts. These properties can be personalised to an end user's preferences and context. They allow users to make semantically meaningful and expertly enriched queries across heterogeneous information sources. A semantic attribute is the key concept of SARA and is significant to steps 1, 2 and 4 (domain understanding, data understanding and personalised query formulation) in the SARA process model. Semantic attributes can be quite simple and atomic and just describe a single characteristic of a domain. But their main value is that they can be highly complex. Then they contain sets of links to domain data, parameters and associated rules defined by experts. Furthermore they can be composed together to enhance their meaningfulness even more. The architecture allows that new semantic attributes can

be incorporated whenever there is demand or when they exhibit a separation from the more static domain model. Along with personalisation some sort of fuzzy semantic can be introduced into the framework which other systems in the relational database area are deficient in [**31**].

These characteristics of SARA are fundamental to my implementation. It allows improving the view of a domain and injecting the necessary semantic facets for matching proper user interpretations. Because of that, and central to my work, personalisation and user modelling exhibit areas that necessitate further focus to support deeper exploration of various sources.

2.4 User Modelling and Personalisation

Another closely related discipline is user modelling. UM is originating in other areas of research as Natural-Language Dialogue Systems, Knowledge Representation, Planning and Plan Recognition or HCI (human-computer interaction). User models are pre-eminently used in Intelligent Interfaces, Active and Passive Help Systems, Adaptive Hypermedia, eLearning and other personalisation frameworks to support adaption, prediction, guidance and profiling of users interacting with them [**32**] [**33**].

These days, there is an improved demand for expressive representation systems for user models und their underlying business logic. Hence, interactive systems need to adapt to their users. There is a big variety of different background, interests and goals, levels of expertise, abilities and preferences. Depending on the field, user initiated and user selected adaptation is commonly not the solution if lack of knowledge, time or motivation prevents optimal adjustments [**1**] [**33**] [**34**]. To deal with these flaws, software systems must make assumptions about the user as long as these deem to be relevant for tailoring the overall behaviour. What follows are two different activities that can take place once the software believes the user model has to be adjusted: It can either change automatically without burdening the user (implicit modelling) or get a user's explicit consent. The latter – computer assisted adaptation – goes along with additional explanations and information and is preferable if the fine tuning is infrequent but significant. Both approaches can also be combined to allow further improvement [**1**].

As such, a user model corresponds to a user's characteristics as computed by the system. It differentiates between qualitative and quantitative interest in certain targets or objects [**35**]. Often this includes a large overhead when creating relevant user models. At the same time a wide ranging model is hard to achieve. Therefore it is essential to only concentrate on these attributes that impinge for the most part and represent interesting features of the individual user. Different machine technique methods for obtaining these assumptions have been discussed in [**36**]. These methods comprise passive/active, user-initiated/automatic, logical/plausible, direct/indirect, and on-line/off-line methods.

Interactions usually involve initial interviews to get background data for assigning an individual user to a stereotype group. This takes place when specific preconditions are met and the system is able to accurately characterise them [**37**]. Depending on the domain and the variety of personal attributes it may be fairly difficult to map these characterisations properly. Certain dialogue actions can be utilised to implicitly improve and refine assumptions and the related model furthermore. This involves a downside: It may take some time to put together a suitable model which also might lack confidence if the user model is entirely based on interactions [**38**].

Scrutability is another important aspect when creating personalised adaptive systems. It gives users greater responsibility and control over various aspects regarding the user model. This helps to increase awareness and better communicates what is happening. It thus forms a tighter rational connection with the user which benefits in improved results [**39**] [**40**].

Essentially, it is important to find the right balance for the two modelling approaches. Explicitly asking the user for participating in a tiring personalisation process upfront or during a session might destroy his "experience" and consequently discourage or even repel him. On the other side, implicit modelling may cause disappointment if the model is not correct, the user feels the system not to be reliable, or not to be in charge.

2.5 Film Recommendation Services

The spotlight of my implementation rests upon films and how a recommendation approach in combination with implicit user modelling besides the Semantic Attribute Reconciliation Architecture Process Model works. A great deal of interest lies in the field of available implementations that offer similar services – commercial and complimentary. Among other things, I will point out primary features and their different realisations. I will start off with the one service probably well known by the reader. Then the discussion will gradually enter more specialised systems that demonstrate fairly different but appealing approaches.

2.5.1 The Internet Movie Database

Widely known as **IMDb** [**3**], the Internet Movie Database not only offers a gigantic repository of facts and figures about pretty any film related subject. This contains cast, directors, editors, budget, release date, genres, poster images, taglines, quotes, trivia, awards, filming locations and many more. The set of metadata they have available is enormous and includes nearly everything you can think of. A vast user base with 57 Million visitors per month populates the forums. This takes care of a continuous flow of new content like opinions, comments and further discussions associated with actors or films.

IMDb does not ask you upfront to input ratings or tell what films you like. Their film recommendation system draws its advantage from user correlated data stored in the system. There are a few entry points to get recommendations:

For one, you get an assortment of linked films on each film page. Furthermore they have a charts section [**41**] that allows the user to browse through lists like “IMDb Top 250”, “IMDb Bottom 100”, “All-Time Box Office”, “Votes by Genre”, “Votes by Decade”, “Votes by Gender” and much more.

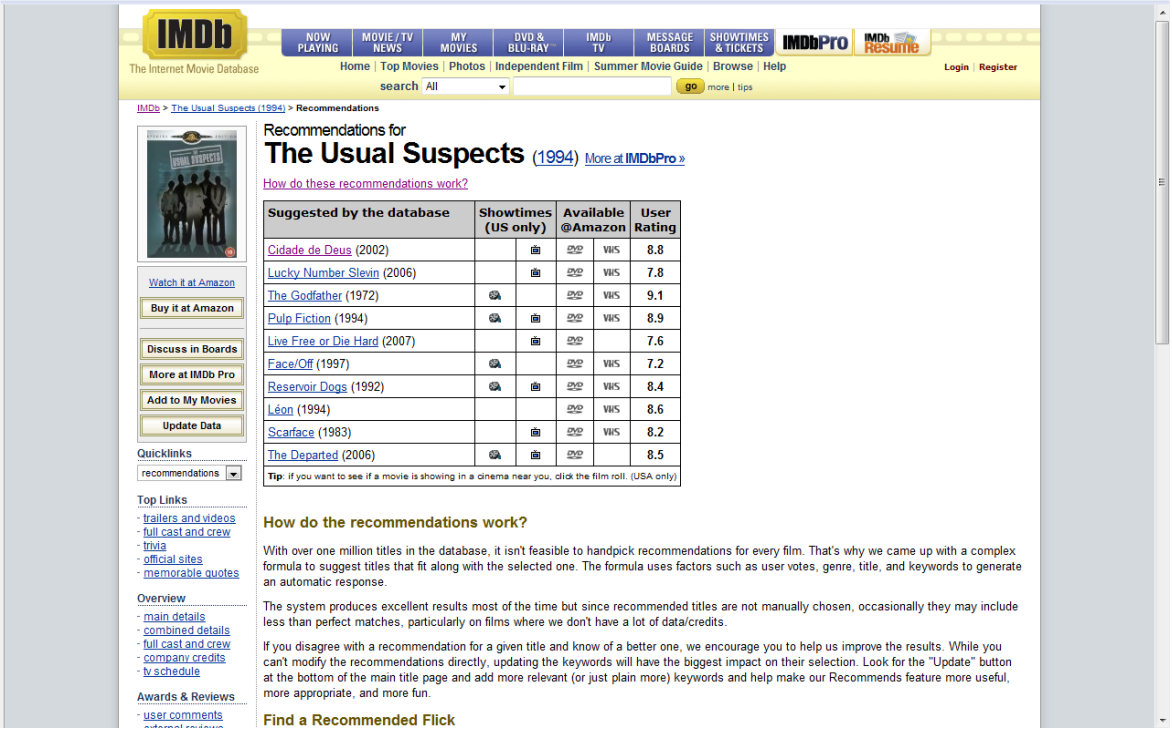


Figure 2: IMDb recommendation page

As depicted in Figure 2 a separate recommendations service site [**42**] allows you to explicitly enter a movie and the system replies with ten film titles it considers closely related. IMDb states about the service:

> *"With over one million titles in the database, it isn't feasible to handpick recommendations for every film. That's why we came up with a complex formula to suggest titles that fit along with the selected one. The formula uses factors such as user votes, genre, title, and keywords to generate an automatic response.*
>
> *The system produces excellent results most of the time but since recommended titles are not manually chosen, occasionally they may include less than perfect matches, particularly on films where we don't have a lot of data/credits."*

Should you not agree with one of the recommendations they kindly ask to update and add keywords since these have the biggest impact on the recommendations feature.

It is apparent that the main features of IMDb are all dependant of the community. Along with some sort of algorithm they are eventually able to make their

assumptions and recommendations. My implementation however does not consume any real time community content relationships.

2.5.2 Rotten Tomatoes

Rotten Tomatoes [5] is a community driven web portal which includes more than 87,000 film titles and more than 200,000 reviews. The site features a "Tomato Picker" which is an interactive film title browser. It allows users to browse through the entire database set by genre, "Tomatometer", MPAA rating, era & decade, and year. The "Tomatometer" shown in the very left column in Figure 3 measures the percentage of "Approved Tomatometer Critics" who recommend a certain film (or the number of good reviews divided by the total number of reviews). A good review is denoted by a "fresh" tomato. A bad review is indicated by a "rotten" tomato. In order for a movie to receive an overall rating of "fresh", the reading on the "Tomatometer" for that movie must be at least 60%. Otherwise, it is "rotten".

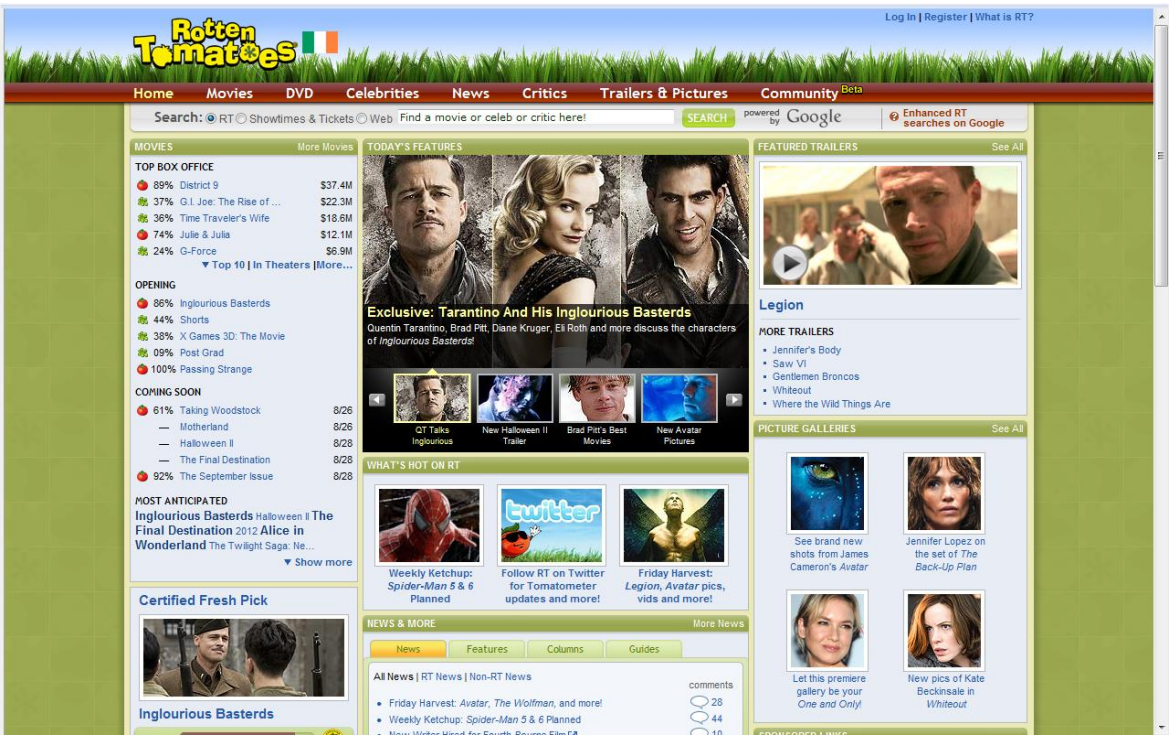


Figure 3: Rotten Tomatoes, Tomatometer

As an alternative of telling Rotten Tomatoes which films you like, you can advise it what type of films you like, which actors you want to see, and other criteria to help it

find the best film for you. There is a lot of variability in the quality of Rotten Tomatoes recommendations but it is also a way to find the right film for any mood.

The system draws its power from its large community that consists of many (semi) professional film critics. Often these critics not only give just a simple rating but also more comprehensive reviews and information about films and related things. With regard to my approach Rotten Tomatoes offers detailed information about films. The "Tomatometer" gives the visitor a brief idea how other users think about a specific title regarding its rating and/or keywords. Also the website offers a section where you can search for films entering keywords, genre, Tomatometer ratings and ranges which is clearly very explicit.

2.5.3 Amazon

Due to its prominent position and primary desire to sell products, I want to briefly outline **Amazon** [**43**]. In first place Amazon bases its films and DVD proposals on information gathered through peer user. The website tracks and collects users while interacting with it. Amazon registers what users buy and relates this type of information back to an item's popularity. This enables some sort of implicit community rating which is furthermore influenced by also other domains as books or music and vice versa. Next to that, users voluntarily create and maintain lists of their favourite film which also aids people that lack ideas. Furthermore, visitors are allowed to rate films and provide additional information should they already own an article or have a strong opinion about it. As previously hinted, Amazon assesses page views. This means, that just a brief interest in a film by visiting the relevant product page may be recorded and will be considered in later suggestions. That way Amazon in combinations with its numerous branches in all major countries is able to build gigantic user models.

Amazon shows a number of interesting approaches and patterns. Their website interface is clearly targeted to transport their business agenda. For that reason it is more functional and less visually appealing, but nonetheless highly functional to display the required information. Amazon's big advantage is their huge user base and their ability to draw from these correlated explicit and implicit interactions which subtly related to this work's user modelling and exploration features.

2.5.4 MovieLens

MovieLens [**4**] is a film recommendation website. You tell it what you love and hate and that information is used to generate personalised recommendations for other films you will like and dislike.

The site is maintained by **GroupLens Research**[**44**] which is part of the Department of Computer Science and Engineering at the University of Minnesota. They conduct research in:

- Recommender system algorithms
- Information filtering
- Interface design
- Online community design and theory
- Mobile computing

As such they use MovieLens as experimental platform to be able to study these areas. The site has thousands of users and these have provided millions of ratings. They use collaborative filtering to generate movie recommendations by matching together users with similar opinions about films. Each member of the system has a vicinity of other like-minded users. Ratings from these neighbours are used to create personalised recommendations for the target user.

All they need to know is your ratings about films you love and hate as exhibited in Figure 4. Once you rate 15 films they can generate personalised recommendations by evaluating your taste based on ratings to films you have seen before. They furthermore offer a feature called “Movie Buddies”. You can be buddies with any other MovieLens user and generate personalized group recommendations. In the QuickPick service a user simply enters one or more films he is interested in and a list of suggestions is returned.

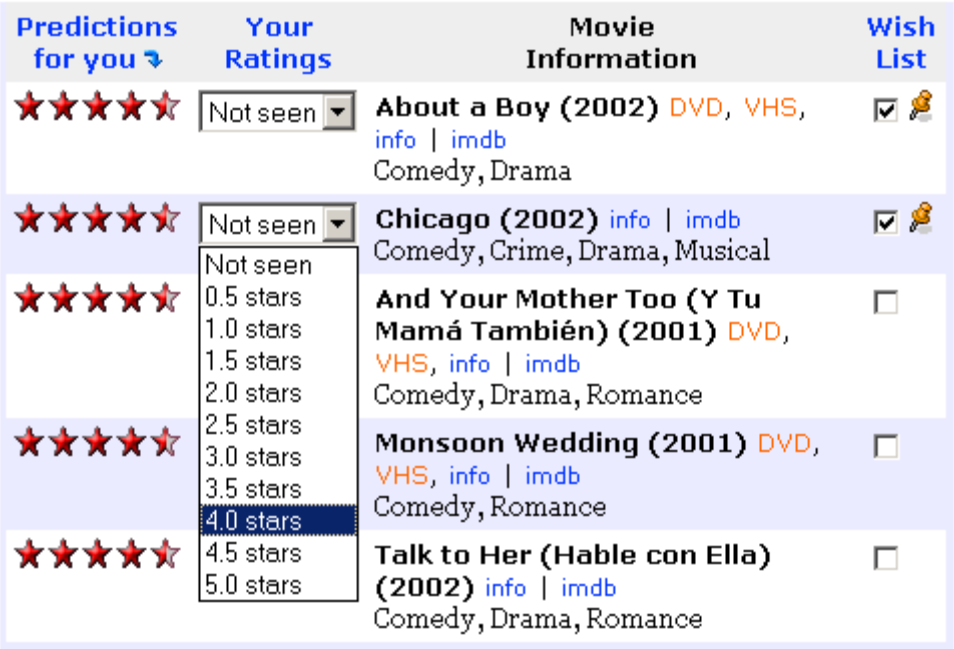


Figure 4: MovieLens recommendation interface

As the owners already state, this system has a highly academic background. As such it makes it rather interesting in regard to my implementation. Especially when looking at their goals and aims. Nonetheless, MovieLens displays a lot of textual information and is not putting that much concentration into graphical appeal and invasive exploration.

2.5.5 What to Rent

Then there is **What to Rent** [45]. Its approach bases on a theory that states, "A movie viewer emotionally interacts with a film in the same manner that he interacts with other human beings."

To exploit this theory, they first interpret a user's personality and form a general model of how he reacts to the world and his average emotional state. This is done by asking 20 questions upon sign up to build a sketch of a user's taste. This profile is then used to deliver film suggestions when demanded.

The interesting fact: The questions asked have almost nothing, or not much to do with films. Instead they are based on topics like relationships and general attitudes, e.g. "How long does it take you to fall asleep at night?", "What kind of experience do

you most look for in your hobbies?", "What percentage of people within 5 years of your age enjoy the same movies that you do?", or "How well can you control your own behaviour?".

Figure 5: What to Rent, Mood assessment page

Once the initial model is processed and stored, the user proceeds to the recommendation screen. He then has to make some last adjustments like informing the system about his current mood and whether he wants to watch something inside or outside his comfort zone. See Figure 5 to get an idea about the mood assessment page. It asks "What type of movie would you like to watch?" to assess what type of evening the user is looking for.

The creators of the service proclaim that there is an inherent problem when trying to decipher a user's personality via a quiz, since most users lie. Users have an idealised vision of who they are and will answer questions in their ideal manner. This leads to wrong responses.

The **What to Rent** personality quiz is designed to put users in situations that they have been in before or can easily picture experiencing. By asking seemingly

unimportant real life occurrences the service expects to get quite accurate responses. This allows relative accurate models of a visitor's personality.

Each movie in the database is evaluated and analysed as if it were a real person. After the software learns what type of mood the user is in, the system performs a few algorithms with the user's personality model, mood, and the films in the database. These operations give an idea of what someone would experience if he watched each movie in the current mood. The film that gives the best experience is subsequently recommended. Additionally, films that already have been suggested, have been seen already, or marked to be ignored are tracked and not considered in the analysis.

With regard to my implementation, **What to Rent** offers an interesting approach by trying to explicitly model a user's mood or state he is in. Even though the unpleasing interface might prevent people from persistently surfing the site, it is functional and results are quite good.

2.5.6 Criticker

As a measure of similarity in films, **Criticker** [**46**] uses a feature named TCI. This stands for "Taste Compatibility Index" and compares a user's ratings (on a percentage) to those of other users and evaluates how closely your tastes match theirs. It combines community and functionality. When matches are found, the implementation will list precisely these people with the most similar taste and then allow you to see their profiles and preferences which span both normal users and published critics. As such it utilises fascinating User Modelling in combination with peer data content.

2.5.7 Clerkdogs

When using **Clerkdogs** [**47**] a user starts off by entering a film he loves and the site returns similar ones. The creators believe that humans give the best movie recommendations and relations. Therefore the suggestions are hand-picked. Their database is made up of hundreds of thousands of individual recommendations from former video-store clerks who are said to understand why customers like films. No algorithms or collaborative filtering is used. Clerkdogs claim they have analysed all

attributes of films to create a database that is much richer and deeper than the collaborative filtering engines.

The website was designed to let customers interact with the database and to take control of their film selection experience. Besides that, they have carefully identified each of a film's top 10 attributes, choosing from among the 36 core attributes (like "action", "bad taste", "complexity", "geek factor", "political" or "so bad it's good") that live on our site. They then prioritised each of these attributes and rated them for quantity on a 1 to 10 scale. Even though this is a number based system, all of the numbers come directly from humans. This system helps to power other applications of the service, including "Mash it" as depicted in Figure 6. It allows driving through Clerkdogs' database by adjusting and changing the attribute sliders to suit tastes. SuperMatch expands the number of selectable attributes to 300. There is no option to rate movies at all.

The core functionality is hence based on content which is added and maintained by expert users which exhibits similarities to my work. Different however is that visitors are able to tweak their user model by explicitly adjusting and weighting certain features to a great extend.

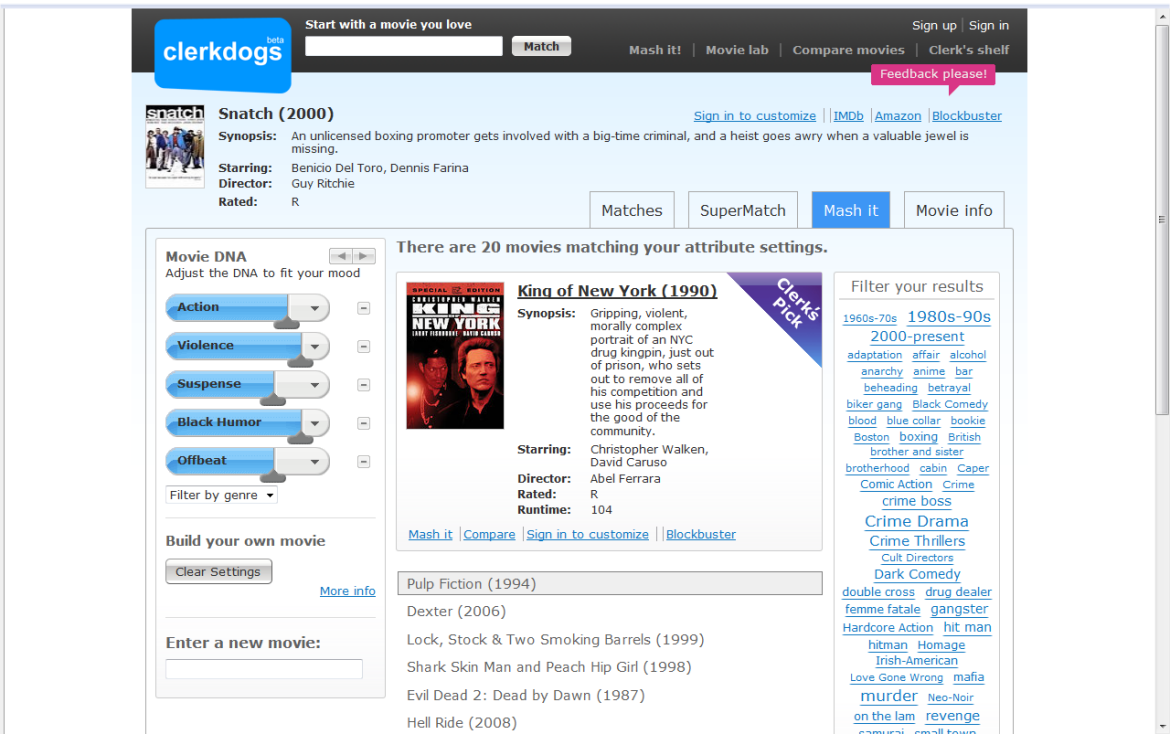


Figure 6: Clerkdogs, "Mash it" app

2.5.8 Jinni

Jinni [**48**] uses a social search-and-recommendation engine titled “Movie Genome”. It is a big project in progress with the Jinni community to describe video more complexly. Using proprietary technology, a Jinni search takes the visitor inside plot elements, atmosphere, and more. Jinni is not a social network. They call it an “internet application designed to fit how people relate to movies”.

Jinni’s recommendations aren't based on statistics, but on aspects of content. The system creates a compound profile of a user’s unique taste and then connects this with films considered to be liked. It combines community and statistics. The user can also use scrutability functionality to find out more why a title appears to be similar to another one. Moreover one can rate films he has seen and get recommendations based on that. Jinni assigns a set of genes to every user which is drawn from aspects of the content he likes. These genes, along with some other instruments, are constantly adjusted as the system learns from ratings, reviews, and other actions on the website.

Figure 7: Jinni

The “Taste Types” look at the characteristic attitudes and preferences that attract viewers to different types of films. Since individuals usually like a range of films, Jinni uses a combination of Taste Types for mapping. It allows the user to search for films based on mood, genres, time period, place, audience, and praise or based on plot preferences. This is shown in the right column in Figure 7. Most interestingly it offers a semantic search where users can input terms like "movies that have gangsters wearing red dresses" or "films where apes rule the world".

As such Jinni is quite comprehensive and merges a lot of different approaches primarily implicit and explicit user modelling.

2.5.9 Summary

There are so many more film recommendation systems out there offering a huge variety of different and very interesting ways to come up with suggestions. To simply name but a few more: Nanocrowd [**49**], TasteKid [**50**], Flixter [**51**], Netflix [**52**], MovieProfiler [**53**], LinkedMDB [**54**], FindAnyFilm [**55**] or Spout [**56**].

Nevertheless, the ones that have been described in more detail above span the State of the Art in this area to a good extend and give a fairly wide-ranging overview of the diverse technologies that are available.

Table 1 below provides some comparisons.

Name of recommend-dation service	**Com-munity based ratings**	**User Modelling**	**Personali-sation**	**Recommen-dations**	**Features**
IMDb	Yes	-	-	Based on private formula	IMDb Top 250, All-Time Box Office, Votes by Genre, specific ten similar films
Rotten Tomatoes	Yes	Explicit	No	Based on user and (semi) professional critics ratings	Tomatometer, keywords, MPAA rating, era & decade, year
Amazon	Yes	Implicit and explicit	Based on peer data and personal data	Based on algorithm	Shows films a user might like

MovieLens	Yes	Explicit	Yes	Automated collaborative filtering	Personalised group recommendations, QuickPick
What to Rent	No	Explicit	20 questions upfront and subsequent mood refinement	Based on some theory, model of emotional state, automated	Personality quiz, mood based
Criticker	Yes	Explicit	No	Based on similar peer data, automated	Taste Compatibility Index
Clerkdogs	No	Explicit	Yes	Handpicked by domain experts	Prioritisation of 36 core attributes , Mash it, SuperMatch
Jinni	Yes	Implicit and explicit	Upfront/on the fly	Based on statistics, but on aspects of content, profile of a user's unique taste	Movie Genome, Taste Types, films based on mood, genres, time period, scrutability functionality

Table 1: Film Recommendation Systems

Chapter 2.6 summarises interesting visualisation techniques that have influenced the design and played a major role with regard to the Film Domain Exploration Client (FDEC) and its design considerations which are discussed in 3.2 and thereafter.

2.6 Visualisation Techniques

Data presentation can be beautiful, elegant and descriptive - even when dealing with large amounts of information. A good visual concept can simplify access and improve results when dealing with vast sets that usually would easily reach a user's cognitive limit [**6**]. There is a variety of ways for visualisation like tables, histograms, pie charts or bar graphs. Most of them are used every day depending on the situation. However, not every single one is applicable for every domain. Sometimes, in early stages of some sort of research, you don't know exactly what you are seeking. As a result, search is more likely an iterative process. It is important to adjust to the circumstances alter techniques to be able to convey results effectively. This means, you don't show all of the results instantly, but present only a subset and adapt to changed requirements throughout the exploration process. There are several

excellent approaches available. A selection relevant to this work will be presented next. Yet, I need to point out that these concepts will not be limited to the film domain and actually take examples from diverse areas.

2.6.1 Cooliris

With Cooliris [**57**] a user is able to browse to large sets of images and incorporates data from Google Image Search [**58**], YouTube [**59**], Flickr [**60**], online retailers and a many more. The interface neatly plugs into the most common web browsers and allows easy access. This is done by explicitly running the add-on or by pointing to a little Cooliris icon that appears on an image on mouse over. After that you see a 3D wall that presents a number of images or even videos related to the one you selected earlier (notice Figure 8: Cooliris on page 27). This interface is fully interactive and reacts to users. One can drag and drop in order to slide along, zoom, play the videos, and click links to jump to other websites. It simply takes advantage of the visual appeal. The system iteratively streams new content and adds that at the borders so the user experience is not interrupted.

There are more features to keep users interested: You can use an inbuilt search dialogue to get images related to a certain topic or simply use predefined channels. These are additional features to keep users exploring. Although it is not required, you can set up an account and have results and search processes stored, shared and personalised.

Figure 8: Cooliris

With relation to my work, Cooliris offers good ideas of how to keep users bound to an interface that allows easy navigation and iterative flow of new data without being interrupted.

2.6.2 Newsmap

Japanese service Newsmap [**61**] uses a treemap to display the continuously changing data pulled from Google News [**62**].

The site shows how to effectively visualise large sets of information on limit screen real estate. By varying sizes and colours, it divides the information to easily distinguishable sub sets as shown in Figure 9 below. These are computed in regard to associated articles. This enables users to rapidly figure underlying patterns and relationships like the importance of an article within its category or in total.

Figure 9: Newsmap

Along with being able to switch certain features on and off, it further uses pop up dialogue boxes for additional information. The visual paradigm offers a quite simple way to deal with overwhelming quantity of information. However, particularly because the interface displays news headlines, there is a lot of text to be shown. Especially with my work's aim the general concepts of this approach is highly important to be more closely examined.

2.6.3 Stacked Graphs

In their "Stacked Graphs – Geometry & Aesthetics" paper Lee Byron and Martin Wattenberg [**63**] discuss a type of complex layered graph that is effective for displaying large sets of data to a mass audience. Although the layered graphs are not interactive, they exhibit great potential. The approach concentrates more techniques for colouring and ordering.

The graphs look organic and pleasing and make the statistical data accessible. The example depicted in Figure 10 shows trends in an individual's music listening as obtained from data in the Last.fm service [**64**][**65**]. The x-axis represents time and each band symbolises an artist. The thickness of a line stands for the number of times

that songs from the artist were listened to in a given time period. Colour saturation is determined by the total number of times an artist is listened, and hue is related to the earliest date at one of the artist's songs were heard.

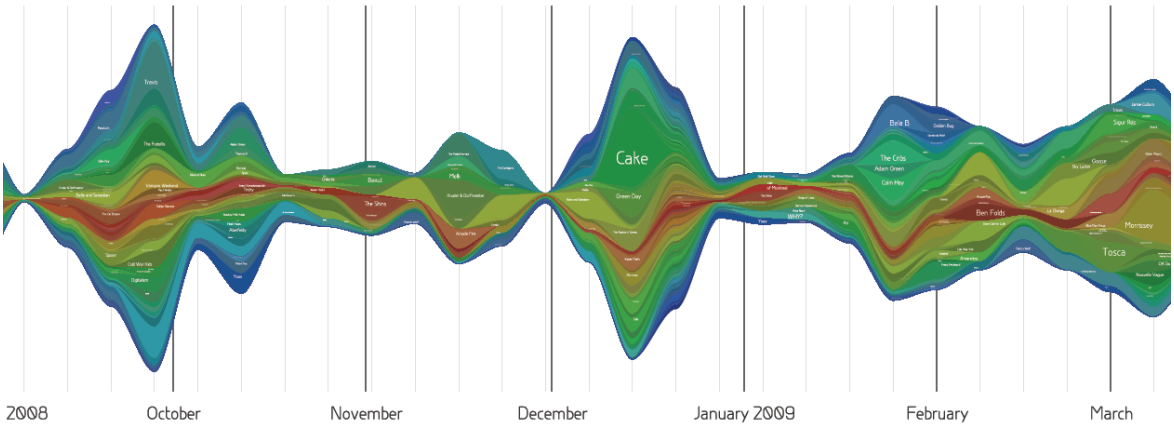


Figure 10: Listening history based on a data pulled from my Last.fm account

What can be drawn in relation to my work is that colours and organic structures make visualisation of any mathematical data more appealing and easier to understand.

2.6.4 Musicovery

Musicovery [**66**] as shown in Figure 11 is a website that displays music taste connections and serves as a customisable internet radio. Moreover it allows users to listen to songs and browse through similar tracks. You can choose and explore music by mood, style, genre, decade and many things more. Filters enable a user to get songs that match positive slow metal or rap from the 80ies. Genres are colour coded and add to the visual appeal of the experience. At the same time it offers a great and simple way to show relationships of highly complex data in a nice fashion. Users have a great number of switches to adjust their preferences and explicitly model their profile. But on the other hand it also allows an uninterrupted experience when browsing the system. This is inherently relevant also to this dissertation's implementation.

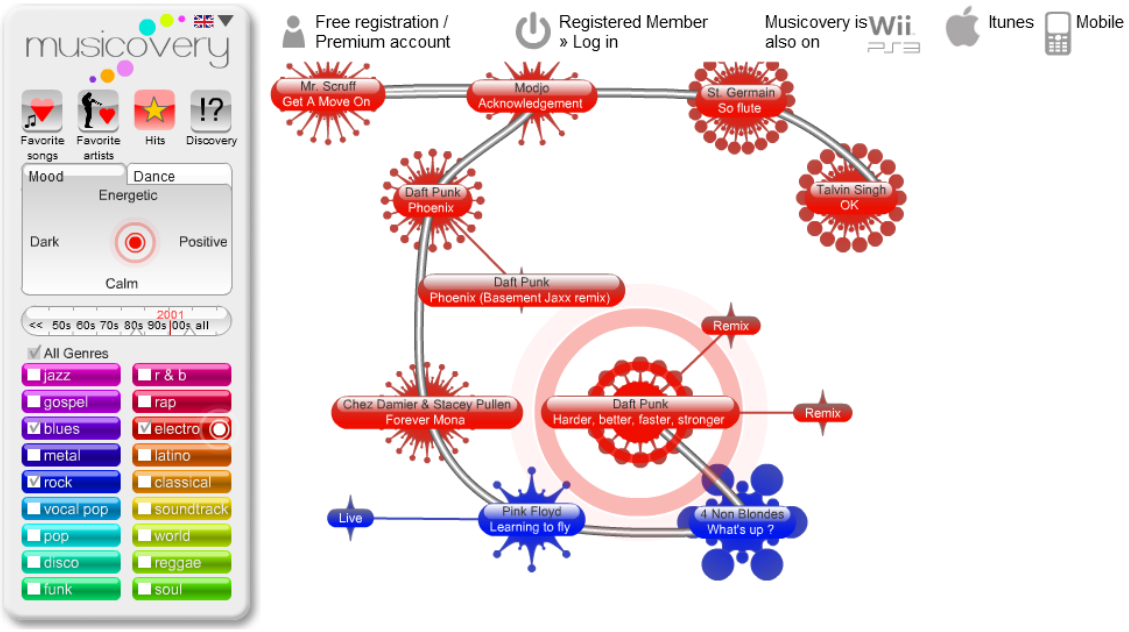


Figure 11: Musicovery: Music taste connections

2.6.5 FIDGT

FIDGT [**67**] is a visualisation software that allows its users to merge various networking sites and interface with Flickr [**60**] or Last.fm [**65**]. Users create tag magnets to be able to demonstrate social network tendencies towards certain things, for example what types of photos are more popular or what genres in music dominate among peers.

Figure 12 shows the interface. It tries to be quite simplistic and engaging at the same time. Objects can be dragged and dropped which results in different weightings among them. The same applies to the stage which makes navigating the system fairly intuitive and fast.

The visual and interaction paradigms have great potential when dealing with a specific number of objects on a stage that might quickly be exchanged, moved around or zoomed.

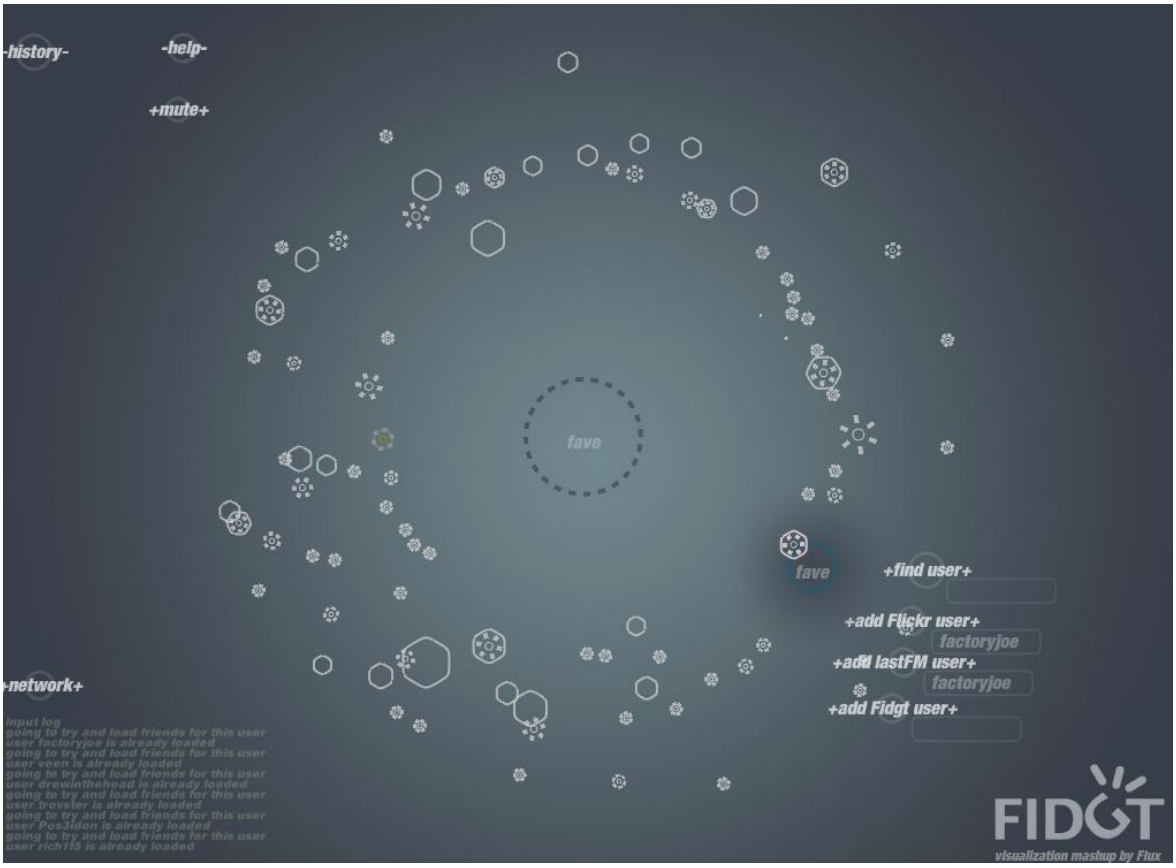


Figure 12: FIDGT - Interface

2.6.6 Summary

Visualisation Techniques are also very comprehensive. The few examples above give an overview about different approaches and how each of these tries to accommodate various user interface, navigation and data visualisation issues. Table 2 shows some side by side comparison.

Name of visualisation technique/ software	**Level of appeal**	**Process of exploration**	**Amount and details of data per step**	**Navigation**	**Personalisa-tion/UM**
Cooliris	High, dynamic, 3D wall	Iterative/ continuous	Huge, basic details (images, videos)	Easy, compelling, non-invasive, drag & drop	Yes, voluntarily
Newsmap	Medium, Treemap	Iterative/ continuous	Huge, lots of text, coloured boxes	Easy, uses switches to enable/disable features	No

Stacked Graphs	High, organic	Static with regard to my example, but allows to be dynamic	Huge, coloured graph with text	Timeline scrolling, items distinguishable by colour, size	No
Musicovery	Medium, map	Constant updates	Medium/low, detailed information	Intuitive to complex, many filters like mood, style, genre, decade	Yes, explicit
FIDGT	Medium	Continuous	Huge	Simple, clean, drag & drop, zoom, tags	No

Table 2: Visualisation Techniques

The Table above leads to a couple of important design decisions which have great impact on visual key features of the Film Domain Exploration Client as will be detailed in 3.2 onwards.

2.7 Conclusion

In this chapter I pointed at a number of key areas that are relevant to this dissertation. I started with KDDM, Flexible Querying and Dataspaces and I further explained their requirements and what they try to achieve. In detail I evaluated different approaches and current implementations. A deficiency in this area and more complex needs lead to the development of the SARA process model and the utilisation of semantic attributes. My implementation and its results depend deeply on both. How these parts of the State of the Art influence the fields of personalisation and user modelling was described in 2.4.

The two sections that follow then show various paradigms of existing film recommendation services and data visualisation approaches. This allows me to draw conclusions for my own implementation regarding implicit and explicit modelling techniques and how UI design supports exploration. Design and implementation follow in the next two chapters.

3 Design

3.1 Overview

In this chapter I lay out what core technologies and design principles have been used to accommodate State of the Art influences and requirements especially with regard to film recommendation services as well as visualisation techniques. The chapter shows where the advancement of my work can be found. In first place this includes considerations of how to successfully combine and balance implicit and explicit user modelling techniques with different subjective expert perspectives. This is to support and motivate users exploring film resources to collect films. As a consequence I am going to discuss aspects relevant to the exploration tool, its visual metaphors and underlying technologies. Furthermore I explain the backend architecture which my client application strongly relies on and what effect these facts have on the design decisions.

3.2 Initial Considerations and influences from State of Art

There are a couple of key aspects that have to be put in consideration. In order to allow people to non-invasively deal with large sets of information, the data has to be presented in a way which reduces cognitive overload. This usually means to apply advanced searching and filtering functions to decrease the amount of information shown at a single time.

There are two ways to handle those filtering parameters. One way is to allow a user to specify what he is looking for. The other one is to have the system decide. The second method is based on some reasoning which calculates what the user should be presented with. Both approaches exhibit pros and cons. They depend on various factors and the projected outcome: Different grades of familiarity with a system result in a changing quality and sophistication of query parameters also depending on the search interface. The user interface plays a major role in what is possible and what is not. Limited features might minimise what expert users can do. But on the other side when it is easier to use, such a system makes itself more attractive to a less experienced audience which can be drawn from the previous chapter.

User modelling techniques and personalisation can be a key concept assisting with this. They help to figure out what users may be like and what they are looking for. The system is able to decrease uncertainty and provide better results. The simpler a user interface is, the bigger the potential audience. Nevertheless, this shifts the amount of explicit modelling techniques towards implicit ones and requires a more refined reasoning and additional expert knowledge to support and enrich the system.

3.3 Requirements

As learned in 2.5 and 2.6 keeping the right balance between a compelling application and adequate usability is important. The UI should be easily accessible from a technological point of view as well when it comes to usability. The application is rich in graphics and emphasises on visual exploration, unfortunately barring visually impaired users for this reason. This is usually neglected to a certain extent with State of the Art recommendation systems.

Since the system primarily learns from interactions with it, the visual metaphors have to be consistent and clear throughout the system. This makes the exploration progress as fluent as possible and enables a continuous experience.

A constant flow of exploration is brought together with pre encoded expert knowledge in form of semantic attributes. These have to be set earlier in the progress by using some specific authoring tool. This tool was developed in connection with the Semantic Attribute Reconciliation Architecture. SARA's underlying process model aims to support users that explore various data sources for valuable information [**13**]. Semantic attributes represent discrete characteristics of a domain preset by experts. These attributes can then dynamically be modified by the client application to fit an end user's preferences and context.

In a continuous progress, users agree or disagree with the data that is shown to them. This allows the system to adjust the user model and refine its suggestions. The implementation should accommodate rapid focus changes.

Figure 13 below shows that joining implicit and explicit user modelling techniques, adjustable semantic attributes, and pleasing visual exploration, prevails over a range of deficiencies that any single one of these would exhibit. Before I am going to talk

about the various technologies that have been used, I will discuss the design of the interface and its correlated visual metaphors.

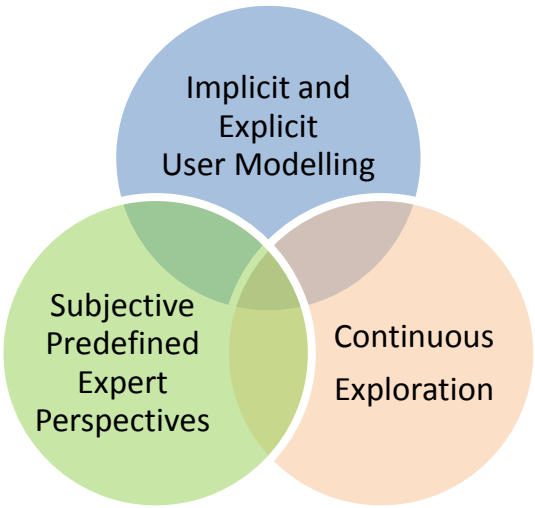


Figure 13: Key concepts

3.4 User Interface and Visual Metaphors

The visual user interface is another important building block. It is the crossing point between the client application and the human user. For this reason it is vital that a UI is able to successfully convey certain types of information. With that I mean that it should be able to be easily understood by the user. [**68**]

Interaction metaphors and patterns need to be as consistent as possible to reduce confusion. Depending on the subject, too much complexity should be avoided. This might distract or exhaust a user and may harm the tool reaching its actual objectives and goals.

Reoccurring interaction allegories that can be associated to real world examples help to mitigate the learning curve. This can be furthermore supported by apparent colour coding as well as by a clear and consistent shape and size of the diverse interface objects. Performance and speed of the application is another important issue that should not be neglected. Especially in dynamic environments where objects change a lot, it is essential to have expected paths and moving patterns of objects. This assists users to understand what is happening. Mixing 2D and 3D metaphors can cause problems in some occasions. It is shown that if these metaphors are reduced to a

minimum, they are comprehensive and understood easily [**69**]. This enables interesting UI decisions.

The Film Domain Exploration Client (FDEC) application is to present film posters. These posters represent real films. Visual representations should be supported by a textual explanation which means that showing titles in addition makes it easier to identify certain films. As it can be seen in Figure 14, posters are arranged orbiting around one central film poster which is placed in the middle of the main stage.

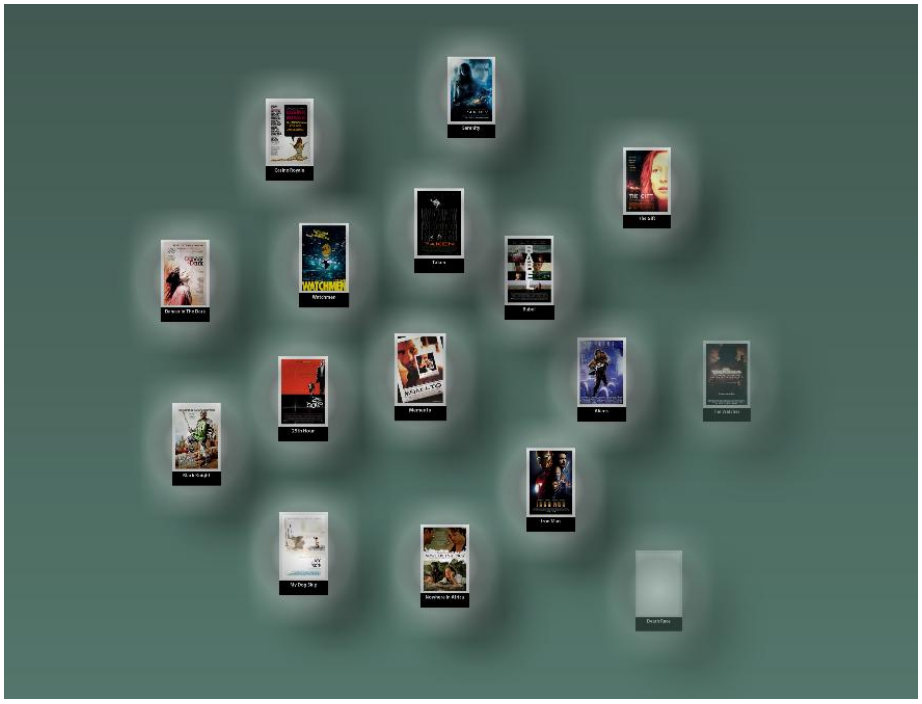


Figure 14: Film posters are placed around a central film

Moreover the interface should encourage exploration. Rating posters and shifting focus refines the sets of items that are presented. A number of 20 to 30 posters at one time has two of the following effects: The user is not overwhelmed with too many objects at once. A number higher than 8 or 10 gives the application enough diversity to grab sufficient attention and support the explorative spirit of the interface [**6**].

3.4.1 General exploration process

Figure 15 briefly sketches the high level interaction patterns and exploration processes within the client. More extended information follows in 3.4.2

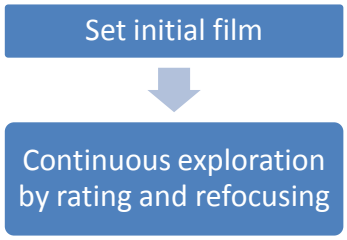


Figure 15: General Exploration Process

3.4.2 Selecting initial film poster

The first step in the procedure of using the application involves selecting an initial film poster. This initial film is then going to be used by the application to calculate and display related posters. The selection screen allows the user to type in a film title as an early sketched version shows in Figure 16 and the current implemented one which is depicted Figure 17. The system assists with auto complete functionality and a drop down list to make selections by navigating with the arrow keys and confirming with enter, or using the mouse.

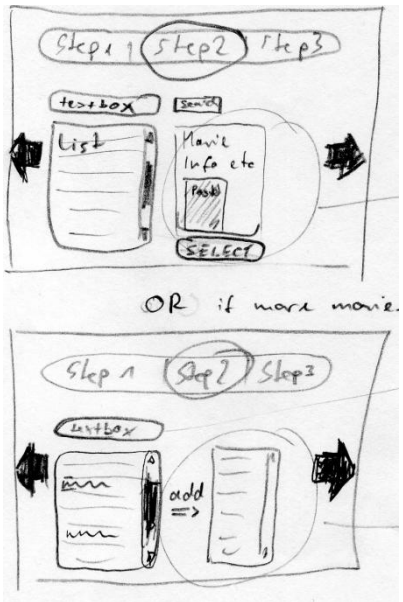


Figure 16: Early sketch of film selection screen

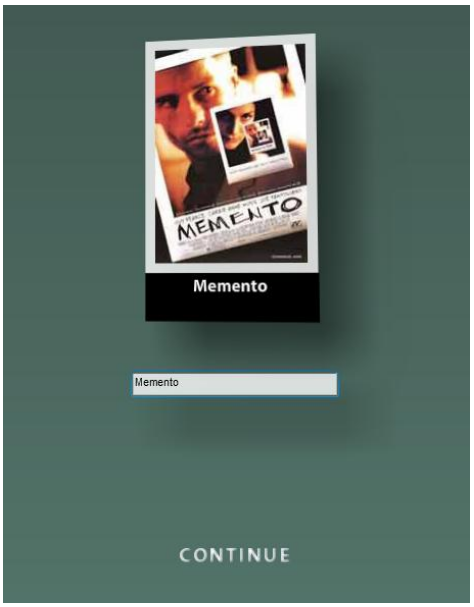


Figure 17: Initial film selection screen

This first step is included to give the user a chance to select his preference film. After he has done that and clicked to continue, the main part of the application comes into focus.

3.4.3 Main exploration

In order to ensure that a user is able to recognize his preference movie, the posters are placed on stage one by one starting in the centre and continuing along a clockwise spiral path as visualised in Figure 18.

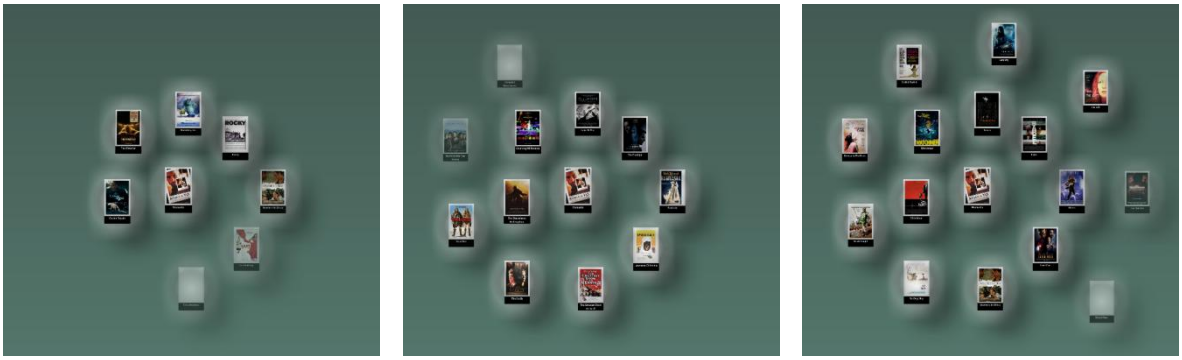


Figure 18: Spiral clockwise placement of film posters

Since one of the application's primary objectives is to allow for exploring films and in order to keep the UI as clean as possible, no other objects will be shown that may disturb the user. Using graphical representations of film posters should further satisfy an enjoyable experience. It has to be underlined that the application is using the posters to represent the corresponding films and not the art work of the posters themselves.

There are just a few simple to understand ways how a user can interact with the system. The main paradigms are as follows:

Zooming: There are several distinct levels of zoom. Using the mouse wheel allows a user to access each level step by step. In strong relation to real world metaphors zooming in or out changes the level of perceived details. Zooming in makes it easier to recognise and interact with certain films whereas being on a higher level gives the user a better idea of the whole set that is currently presented. See Figure 19. This permits to assess films or areas that haven't been looked at or rated so far.

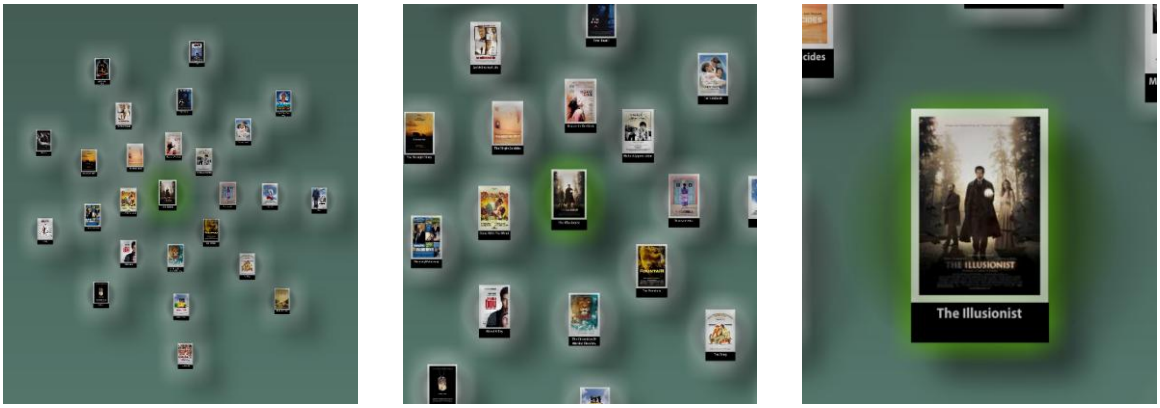


Figure 19: Different zoom levels

Rating: Rating a film poster is key to the success of the system. It draws a lot of parameters for the reasoning within the personalisation and modelling component. Hence, the rating metaphor should be easy to comprehend and also be consistent. There are three different degrees a rating can take: *Like*, *dislike*, *neutral*.

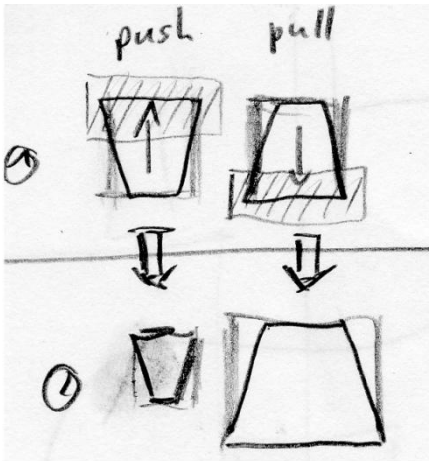


Figure 20: Early sketch of push/pull metaphor

Every poster starts in the neutral phase and – to obey some metaphor - can be *pushed away* to dislike by clicking the upper or *pulled closer* to like by clicking the lower part (Figure 20).

This adjusts the level of preference as depicted in Figure 21 and Figure 22. Arrows indicate the direction in 3D space.

Figure 21: Rate film – “Like”

Figure 22: Rate film – “Dislike”

To assist categorising rated films for the user, some contrasting colour coding is introduced. Films that have been liked, exhibit a faint green glow and appear closer (Figure 23). Neutral films are neutral (Figure 24) and red indicates disliked ones (Figure 25).

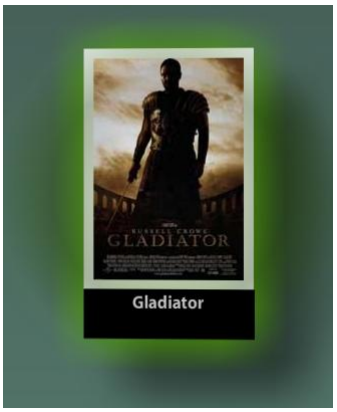


Figure 23: Liked film

Figure 24: Neutral film

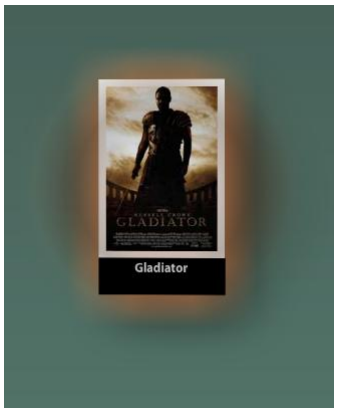


Figure 25: Disliked film

However, should a poster that has been rated previously within the current session come up again, the rating is restored and the poster will be shown with the according visual adjustments. This makes it easy to recognise any previous actions.

Collecting: In addition to that, rating a film also stores it in one of 3 different and constantly maintained and updated collections. A so called collection comprises all liked films (as can be seen in Figure 26 below). A different one has neutrally rated films and the last one contains films that have been disliked. These sets are always accessible by zooming out a certain number of levels as of conforming to the metaphor of *getting to higher ground* and consequently gaining some sort of historical overview.

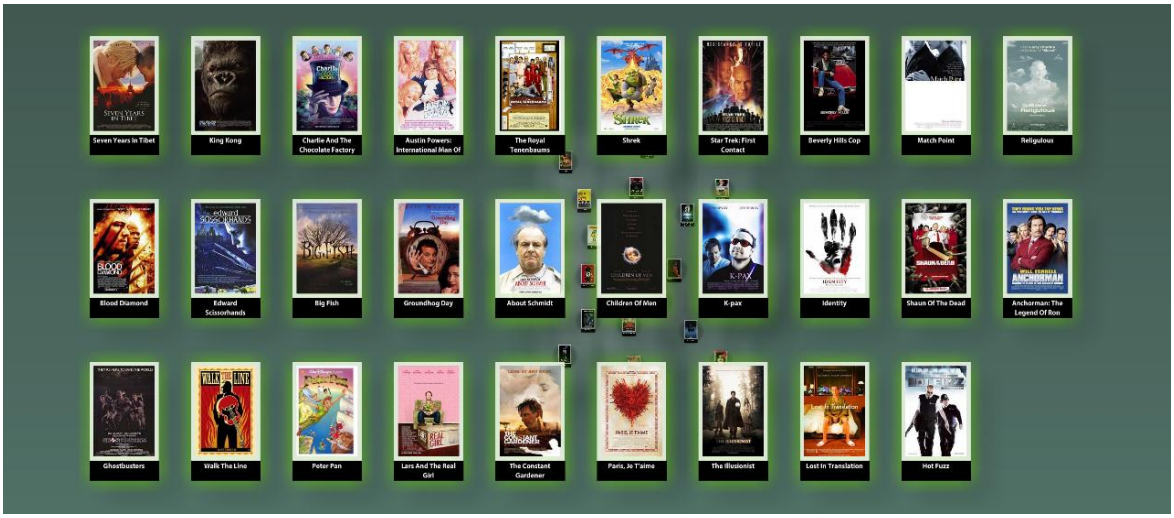


Figure 26: Collection of liked films

Refocusing: As shown in Figure 27 a new focus film can be easily selected by clicking an appropriately labelled button on every film poster.

Figure 27: Set As New Focus Film

This causes some major changes within the application resulting in shifting the clicked poster to the centre, and determining and placing a new assortment of orbiting films. Replacing the images is supposed to feel rather organic and non invasive. It should be clear what happened and that new films have been put on stage.

Panning: Dragging and dropping on the background regardless of the zoom level makes it possible to navigate horizontally and vertically. A user can reach areas that show films not having been rated so far without leaving the current zoom state. This enables a quick and constant navigation and supports the explorative character of the application.

A number of early sketches and mock up design that were part of developing the user interface and interaction metaphors can be found in Appendix C: Sketches.

3.5 Technologies

Following the requirements and decisions regarding the user interface and visualisation metaphors, I am now going to describe the numerous technologies that have been used in the project. In particular I point out the reasons that lead to choosing certain technologies.

3.5.1 Backend Architecture

Although my client application along with its user interface described above is the visual frontend of the system, I describe the backend first. This is, because the backend architecture necessitates and heavily influenced a couple of very important technical design decisions.

The SARA framework provides specific operations to client applications. It enables personalisation of a user's request by using subjective perspectives of domain experts. These personalisable semantic attributes are discreet encodings of domain expertise and offer a subjective way to use domain knowledge other than collaborative filtering or crowd surfing. These attributes have to be incorporated by a specific authoring tool and its GUI [**13**].

The Semantic Attribute Reconciliation Architecture is implemented using Java and Apache Tomcat server technologies [**71**] and further offers an Application Programming Interface (API) which makes it being accessible by client applications.

The actual domain data is stored in an attached eXist database [**72**]. eXist is an open source database management system using XML technology. It stores XML data according to the XML data model and features efficient, index-based XQuery processing [**73**].

For this reason all domain specific data that has been gathered had to be processed and collated to fit into eXist.

3.5.2 Frontend Architecture

The Film Domain Exploration Client (FDEC) was developed using Adobe Flash Platform Technologies [**74**]. There are a couple of reasons at hand:

Flash allows developing highly graphical interfaces and as such very compelling applications. Java Swing briefly came into mind, but the author's familiarity with ActionScript3, Flash's programming language, small file size and the extensive distribution of the Flash plugin (of more than 99%) within browsers supported to go with Flash. At one stage of the development process it came apparent that the backend uses BlazeDS [**75**]. BlazeDS is server-based Java remoting and web messaging technology. It allows clients to connect to distributed data and also push data in real time to Adobe Flex and Adobe AIR applications. The Flex framework [**76**] is a cross platform, open source framework and served as a development environment. It is used to create Rich Internet Applications (RIA) that run identically in all major browsers and operating systems as well as on the desktop with Adobe AIR [**77**].

Essentially, it would not make any difference but I decided to create an AIR application to run on a user's desktop and be able to connect to SARA's API.

For testing and evaluation reasons the client contains some logging (File I/O) functionality which writes session specific user modelling and interaction results to a text file on the user's system.

3.6 High Level Design / System Architecture

This section will describe the overall architecture and operate as a summary of all relevant system components. Figure 28 further includes the common event flow within the Film Domain Exploration Client and also when interacting with a simplified depiction of SARA.

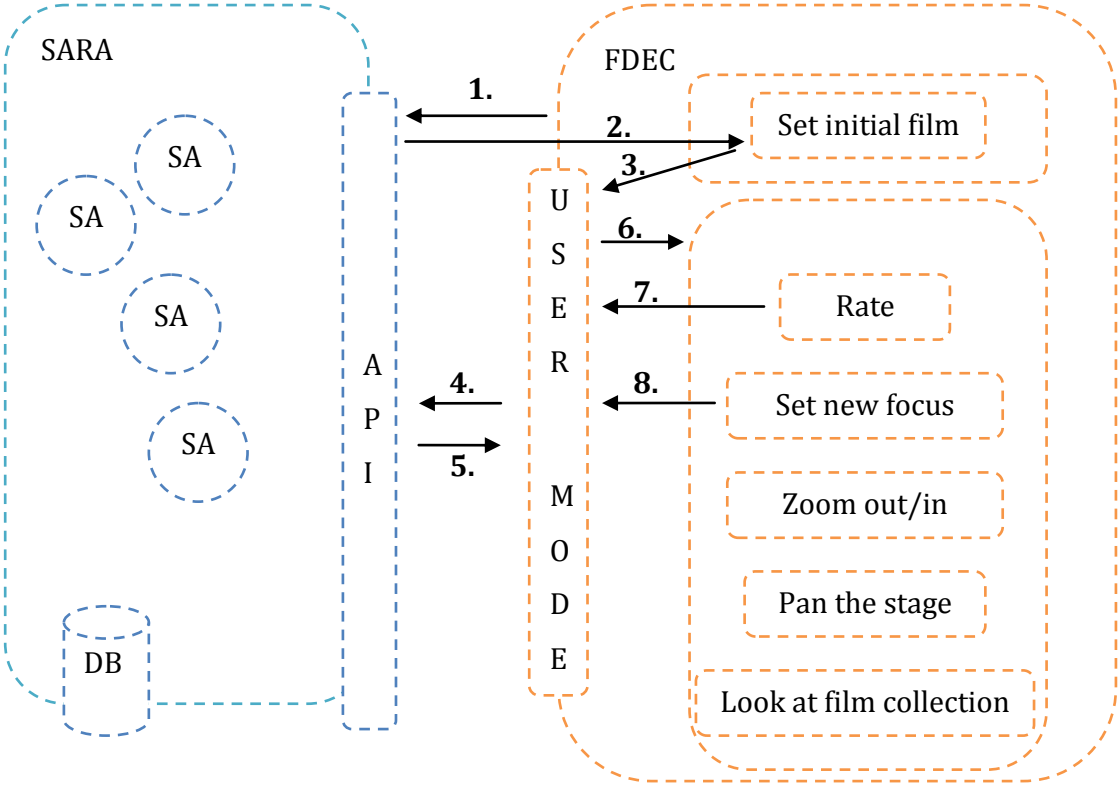


Figure 28: System Architecture and Event Flow

The numbered steps below correspond to identically labelled steps within Figure 28:

1. When FDEC starts, it queries SARA for a full set of all the films in its database.
2. SARA replies and FDEC populates the initial film list.
3. After the user selects an initial film the User Model is updated.
4. FDEC queries SARA with specific parameters for films based on the selected focus film and number of films per semantic attribute based on the UM.
5. SARA replies with as many results as possible.
6. UM places returned film posters on stage.
7. User rates films and UM is updated.
8. Setting new focus on a film asks the UM to update and query SARA

3.7 Semantic Attributes and User Model

Subjective expert knowledge is represented through semantic attributes. These are the key to the modelling approach and they need to be balanced. For this reason the exploration tool should be able to adjust the weightings. It should further leverage a short tail which represents the overall exploration history to deal with temporal shifting satisfactorily.

The client application is able to handle numerous different semantic attributes that can be added and removed on the fly. The total number of film posters that are shown on stage would define the share a single semantic attribute gets.

In my application there are currently 4 different semantic attributes in use. They are all based along the popularity axis and exhibit a different subjective flavour of what people might portray as popular. The attributes are:

- Popularity Rating as rated on IMDb.
- Worldwide Grossing is based on box office results.
- Award Winners as of Academy, BAFTA, Cannes, Berlin, MTV...
- Ratio between Worldwide Grossing and Budget.

By rating films that have been shown with regard to a specific semantic attribute the user implicitly regulates the underlying balance and changes the User Model which has been assigned so far. A different balance has a strong effect on the number of items per Semantic Attribute.

On the other side, the focus film defines the query parameters that are sent to SARA. They define a certain perimeter. When responding, SARA returns films that are within these boundaries.

In the end a user will face different films after setting a new focus film based on his most recent preferences. However, this approach still allows coming up with films that have been rated previously that at are similar based on a different aspect of the popularity axis.

3.8 Conclusion

I have discussed all the issues that are necessary for the accomplishment of this project. The reader got an idea why incorporating abstract subjective expert knowledge by using semantic attributes helps modelling and guiding a user's intent. It shows how deficiencies in current implementations can be overcome. A simple but compelling visual user interface supports the users' urge for exploration during. Certain interactions with the application are used to gather information to support implicit and explicit modelling methods. The system should be able to quickly adjust to changed preferences but also maintain some short history to allow a user to break out of temporal shifting. In the next chapter I will detail about the technological implementation features and how the system has been built.

4 Implementation

4.1 Overview

Based on the design decisions that have been detailed in chapter 3 I will discuss the software implementation of the Film Domain Exploration Client (FDEC) below. In particular the various technologies that have been used in the overall development will be described. The chapter further includes approaches and specific challenges that have been faced while trying to satisfy primary considerations, objectives and goals.

4.2 Technologies

The project comprises a number of different hardware components that have been used. The main development platform was a Dell XPS M1530 laptop with an Intel Core 2 Duo T8300 processor and 4GB of RAM running Windows Vista Home Premium edition. For extensive data processing a remote AMD Athlon 64 X2 Dual Core 6400+, 8GB of RAM, running Windows XP was used at various stages of the project.

Compared to the limited amount of involved hardware, the software was more comprehensive. The list below contains IDEs, programming languages as well as technologies in general: Adobe Flash CS4 4 Professional [**74**], Adobe Flex Builder 3 [**76**], Flex 3 SDK [**76**], Actionscript 3 [**78**], BlazeDS turnkey 3.2.0.3978 [**75**], Altova XMLSpy 8 [**79**], eXist 1.2.6 [**72**], XML 1.0, XQuery 1.0, XPath 2.0, Eclipse Ganymede and Galileo [**80**], Java 1.6, SAX [**81**], DOM [**82**], Apache Tomcat 6.0.14 [**71**] and Adobe Photoshop CS4 [**83**].

4.3 General Implementation History

4.3.1 Initial considerations

In the early stages of the project a domain had to be chosen. It was important that this domain fits well to the project's goal which has been summarised by that time as:

> *"This work will investigate the challenges of modelling a user's intent as he interactively explores a knowledge space. As this exploration is visual in nature, the user will not be burdened with unnecessary modelling and information seeking concerns in order to maintain flow and immersion. As such, a balance between explicit and implicit modelling techniques needs to be maintained, in order to ensure the user has a contiguous experience. The personalised nature of the interactions and their subsequent discoveries presents a need for detailed, yet potentially rapidly changing user models."*

As a consequence the film domain was chosen. Although it can be classified as a so called "soft" domain, it offers different clear advantages.

For one, it seemed that meta data that is essential to the project would be available from various sources and it would be fairly easy to obtain. The visual component of the system was supposed to be very important. Using videos or trailers was not feasible at all. However, film posters do a pretty good job representing real films. This was a perfect solution in order to incorporate a domain where plenty of meta and image data is freely available. That way, both the data mining and the visual element could be ticked off. Another important motivation towards picking films was that many people know a lot more about this media domain and are rather interested in it. This superior interest improves attraction.

4.3.2 Data sources

Semantically linked data is not commonly available. For that reason the next step in the process was to get adequate data. I found numerous resources, websites and/or web services that offer relevant data sets. Apart from actually getting it, a big effort was spent to check licensing terms and copyright issues.

With regard to how much communication is necessary and whether we can consistently query dynamic sources, it was briefly considered to utilise real time APIs of some of the services.

The decision against these was made in order to be independent from congestion or reaching daily quota limits on certain APIs that allow only a mere 1000 queries a day. Caching would have been a solution, but was often prohibited by the terms of service.

Table 3 gives an overview of potential data sources. In the end several different approaches had been taken to obtain the data.

Name	**Available meta content/ primary**	**File format**	**Files obtained**	**Used in project**	**Harvesting approach**
IMDb [**42**]	Filmographies of cast and crew members, genres, keywords, ratings	TXT, CSV, LIST	Yes	Yes	Download
Freebase [**84**]	Genre, actor, producer, writer, title, year	TXT, CSV	Yes	Yes	Download
NY Times [**85**]	Critics picks, ranking lists, 1000 best, best pictures/awards	XML	Yes	No	API
Infoplease [**86**]	Film festivals, movie awards	HTML	Yes	Yes	Parsing
MovieLens [**4**]	Ratings, tags, user data	TXT, CSV, DAT	Yes	No	Download
Netflix [**52**]	Ratings, rental stats	XML	No	No	-
Box Office Mojo [**87**]	Box office data, ratings, budget, grossing, theatres,	HTML	Yes	Yes	Parsing
Wikipedia [**88**]	Film festivals, film awards	HTML	Yes	Yes	Download
The Numbers [**89**]	Box office data, budget, grossing, advertising, sales	HTML	Yes	Yes	Parsing
Allmovie [**90**]	Genre, time, ratings, year	-	No	No	-
Rotten Tomatoes [**5**]	User/critics ratings, Tomatometer	TXT, just small subset available	No	No	-

Table 3: Film resources

This exhibited enough diversity to form a rich basis to work with. The sources I actually gathered data from are also marked in Table 3. There are still a lot more sources available on the internet. However, it was essential to concentrate on a selected group and continue to work with the data at hand.

A big issue was the amount of acquired data and the fact that most of it was not formatted consistently. Handling gigabytes of unstructured text files can create some unplanned problems. Just to give a few: The MovieLens data sets exhibited a total of 737 MB, Freebase film related files were 218 MB and the IMDb movie domain for example consisted of 1.25GB of plain text files with tens of millions of lines.

To handle different sources and file sizes in that magnitude, various text tools are available. In my case I primarily used VIM 7.2 [**91**], Notepad++ 5.0.3 [**92**], SlickEdit 2009 14.0.1.2 [**93**] and Windows Grep 2.3 [**94**]. To get data from sites that did not offer any API or free download, it was necessary to harvest the information right from their website. For this reason I created a couple of tools in Java leveraging SAX and DOM to be able to extract what is needed from relevant HTML pages.

In addition to that, I accumulated a number of plain text files (TXT, LIST, DAT) – structured and unstructured ones, as well as comma separated files. Especially IMDb and Freebase offer their sets in these formats. Since these they contain very valuable information, it was imperative to get them into some accessible format. Once again a few Java tools were developed to parse those divergent files line by line and build proper XML representations.

Eventually I ended up with an assortment of manageable XML files and a huge number of data. Nevertheless, most information was not complete. Some sets still contained wrongly formed tags or simply did not reach the level of quality to be of further use. With the help of XPath and XQuery, eXist and Altova XMLSpy I was able to deal with these downsides. It took a few iterations and many hours of rendering to have all necessary queries of the types shown in Figure 29 processed. Doing mathematical comparisons (to check for films that are tagged with a wrong year for example), merging and removing duplicates with XQuery over hundreds of thousands of lines in multiple XML files concurrently can be fairly time and resource consuming.

After collating films from all different sources I ended up with approximately 1,800 films to form a complete set.

```
<root>
{
for $m in doc("step4.xml")//movie
let $o := doc("imdb_ratings_main.xml")//movie
let $n := doc("freebase_film2.xml")//movie

return
        <movie>
                <title>{ data($m/title) }</title>
                <url>{ data($m/url) }</url>
                <rating_imdb>{ data($o[lower-case(title)=lower-case($m/title) and
                        xs:int(year)>=xs:int($m/year)-1 and xs:int(year)< xs:int($m/year)+1 ]/rating)
                        }</rating_imdb>
                <year>{ data($m/year) }</year>

                <runtime>{ data($n[lower-case(title)=lower-case($m/title) and
                        xs:int(year)>=xs:int($m/year)-1 and xs:int(year)< xs:int($m/year)+1 ]/runtime)
                        }</runtime>

                <budget>{ data($m/budget) }</budget>
                <grossing>{ data($m/grossing) }</grossing>
                <grossing_worldwide>{ data($m/grossing_worldwide) }</grossing_worldwide>


                {
                $m/award_academy
                }
                {
                $m/award_bafta
                }

                ...

                {
                for $c in $n[lower-case(title)=lower-case($m/title) and xs:int(year)>=xs:int($m/year)-1 and
                        xs:int(year)< xs:int($m/year)+1 ]

                return
                $c/cinematographer
                }

                ...

                {
                for $g in $n
                where title=$m/title
                return
                $g/genre
                }

        </movie>
}
</root>
```

Figure 29: Sample XQuery which was used to collate data sets

The main tags that have been chosen to be relevant for a film are: title, poster URL (which is used to identify local poster image), IMDb rating, year, duration (runtime in minutes), budget, grossing, worldwide grossing, award won (includes name or event: Academy Award, BAFTA, Berlin, Cannes, European Film Award, Golden Globe, MTV Movie Award, Slamdance, Sundance, Toronto, Venice), director, producer, cinematographer, actor and genre. Figure 30 below shows an XML example of a single movie. It has to be noted that there are still films that do lack one or the other tag.

```
<movie>
        <title>Lost In Translation</title>
        <url>lostintranslation</url>
        <rating_imdb>7.9</rating_imdb>
        <year>2003</year>
        <runtime>102</runtime>
        <budget>4000000</budget>
        <grossing>44585453</grossing>
        <grossing_worldwide>106454000</grossing_worldwide>
        <award_academy>false</award_academy>
        <award_bafta>false</award_bafta>
        <award_berlin>false</award_berlin>
        <award_cannes>false</award_cannes>
        <award_europeanfilmaward>false</award_europeanfilmaward>
        <award_goldenglobe>false</award_goldenglobe>
        <award_mtv>false</award_mtv>
        <award_slamdance>false</award_slamdance>
        <award_sundance>false</award_sundance>
        <award_toronto>false</award_toronto>
        <award_venice>false</award_venice>
        <director>Sofia Coppola</director>
        <producer>Ross Katz</producer>
        <producer>Sofia Coppola</producer>
        <cinematographer>Lance Acord</cinematographer>
        <actor>Giovanni Ribisi</actor>
        <actor>Bill Murray</actor>
        <actor>Scarlett Johansson</actor>
        <actor>Anna Faris</actor>
        <genre>Romantic comedy</genre>
        <genre>Drama</genre>
</movie>
```

Figure 30: Sample XML for gathered movie data

Poster images where grabbed from Box Office Mojo. However, not all images had been available through that service. It is essential to provide a uniform user experience to be as complement as possible. Thus, about 120 remaining poster images had to be acquired manually from sites like MoviePosterDB [**95**] or IMP Awards [**96**]. This was done after evaluating the time/effort ratio. The images have

been also processed to be of reasonable size (150 x 222 pixels) to support performance and consistency within the running application.

4.3.3 Semantic Attributes

Having brought together all the necessary data, it was time to create some semantic attributes. Over the time, reasonably different types had been discussed: Clever, cult, enjoyable, successful or popular. Mapping available meta data to build a certain semantic attribute can be rather difficult to do. With the tags that I was able to obtain, going for popularity, was best.

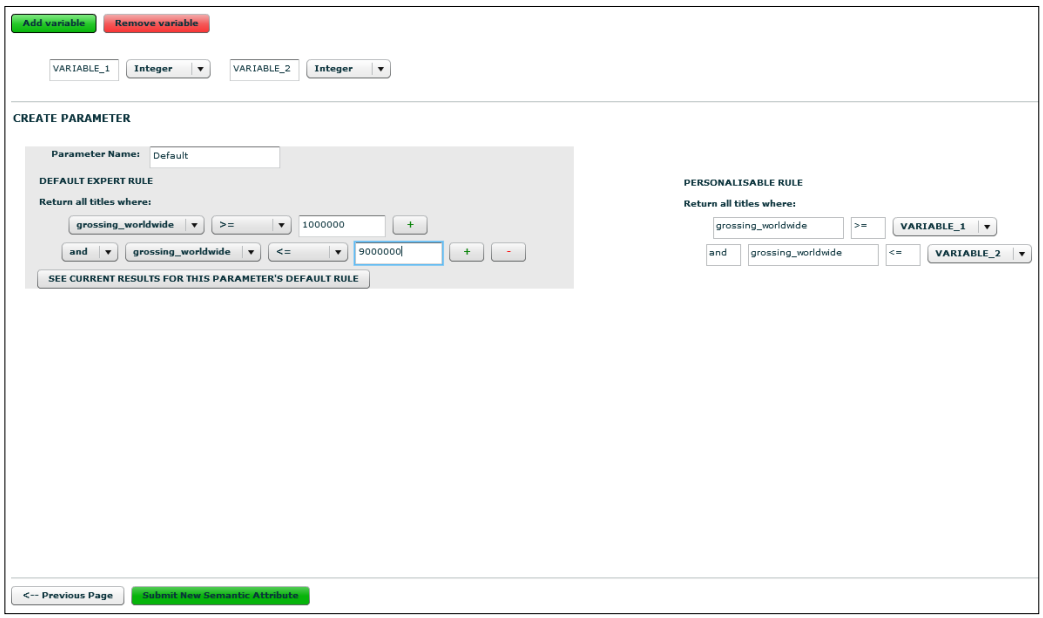


Figure 31: Semantic Attribute authoring tool GUI

With the help of the SARA Authoring tool (Figure 31) I was able to describe four semantic attributes along the popularity axis. My subjective (expert) definitions of popularity can be based on the following different factors:

Rating - Has a film been awarded high ratings?

Grossing - Has it been grossing well?

Awards - Has it won film awards?

Grossing/Budget - How much money did the film make in comparison to its budget?

SARA is a key part, but it is not the aim of this project to explain it and its tools. I am only going to briefly outline what the technical structure of a semantic attribute is and how it looks like. For this reason I will describe a personalisable semantic attribute called "*PopularGrossing*" (Figure 32).

During the development phase a number of additional attributes have been created and used. They allow querying for similar films based on different axes and involve cinematographer, director, producer, actor, genre or year. However they are disabled in the current version to concentrate on the popularity axis.

```
<?xml version='1.0' encoding='utf-8'?>
<SemanticAttribute>
	<Name>PopularGrossing</Name>
	<SetName>SetPopularGrossing</SetName>
	<Description></Description>
	<TypeOfSemAtt>Instance</TypeOfSemAtt>
	<ComponentMetaData>
		<MetaData>
			<LocationOfSourceFiles>Films</LocationOfSourceFiles>
			<SearchElement>movie</SearchElement>
			<MetaTag>grossing</MetaTag>
		</MetaData>
	</ComponentMetaData>
	<PersonalisableSemAtt>True</PersonalisableSemAtt>
	<ReturnEntity>title</ReturnEntity>
	<PersonalisedVariables>
		<PersonalisedVariable>VARIABLE_1</PersonalisedVariable>
		<PersonalisedVariableType>Double</PersonalisedVariableType>
		<PersonalisedVariable>VARIABLE_2</PersonalisedVariable>
		<PersonalisedVariableType>Double</PersonalisedVariableType>
	</PersonalisedVariables>
	<Parameters>
		<Parameter>
			<Name>Default</Name>
			<Rule> grossing  &gt;=  0.1  and   grossing  &lt;=  1000000000 </Rule>
			<PersonalisedRule> $doc/grossing  &gt;=  " + VARIABLE_1.doubleValue() + " and
				grossing  &lt;=  " + VARIABLE_2.doubleValue() + "</PersonalisedRule>
		</Parameter>
	</Parameters>
</SemanticAttribute>
```

Figure 32: PopularGrossing semantic attribute XML

The semantic attribute outlined above has been created with an Authoring GUI tool which was provided to me and has been shown in Figure 31 earlier. The tool allows you to intuitively encode a subjective perception of any type. In my case this was grossing. Later on in the section about the Film Domain Exploration Client more information regarding how semantic attributes are leveraged, and how the

parameters that define their thresholds can be dynamically customised in a meaningful manner.

For more information I refer to Cormac Hampson's PhD thesis and his paper about "Supporting Personalized Information Exploration through Subjective Expert-created Semantic Attributes" [**31**].

4.4 Film Domain Exploration Client

4.4.1 Goal

This is just a brief recap of what the tool is supposed to convey. Its name already contains a couple of key concepts: The application should let a user explore the film domain and build some sort of a wish list. Criteria for the recommendations are based on a user model and personalisation techniques to balance results. Another feature that has to be put into consideration is the subjective expertise which is encoded through semantic attributes. Combining all of them is the aim of the Film Domain Exploration Client. This is achieved by exploiting and balancing the relation between user model and semantic attributes. Users should find films they know and they may be interested in. The system builds a map to keep track of what a user had been watching and eventually educates the user of what he likes and what he does not.

4.4.2 Structure/Architecture

Next I will describe the high level structure of the application. It includes the main components and how these communicate internally and externally as can be seen in Figure 33.

In essence there are 4 core parts: The stage component deals with the user interface. It contains the poster images that are presented and also handles the closely related film objects. A poster loader component allows loading poster images from the file system dynamically (**5.**).

The UI/Stage component reports all user actions to the User Modelling component (**1.**). This is the central part of the system. It includes the Reasoning Engine which

takes care of modelling the user and further balances what semantic attributes will get what share on screen.

In order to interface with the Semantic Attribute Reconciliation Architecture (SARA), the SARA communication module is used. It contains the logic to dynamically query and adjust the parameters of the used semantic attributes. Internally it talks to the User Modelling Component (**2.**) and externally it interfaces with the backend SARA (**3.**) which connects to an eXist XML database to locate and grab applicable data set based on query parameters that are set by the SARA communication module.

A logging module which sits among all major elements of the systems that have been mentioned above helps monitoring and measuring what is going on internally (**4.**). It is able to output log files to the file system (**6.**).

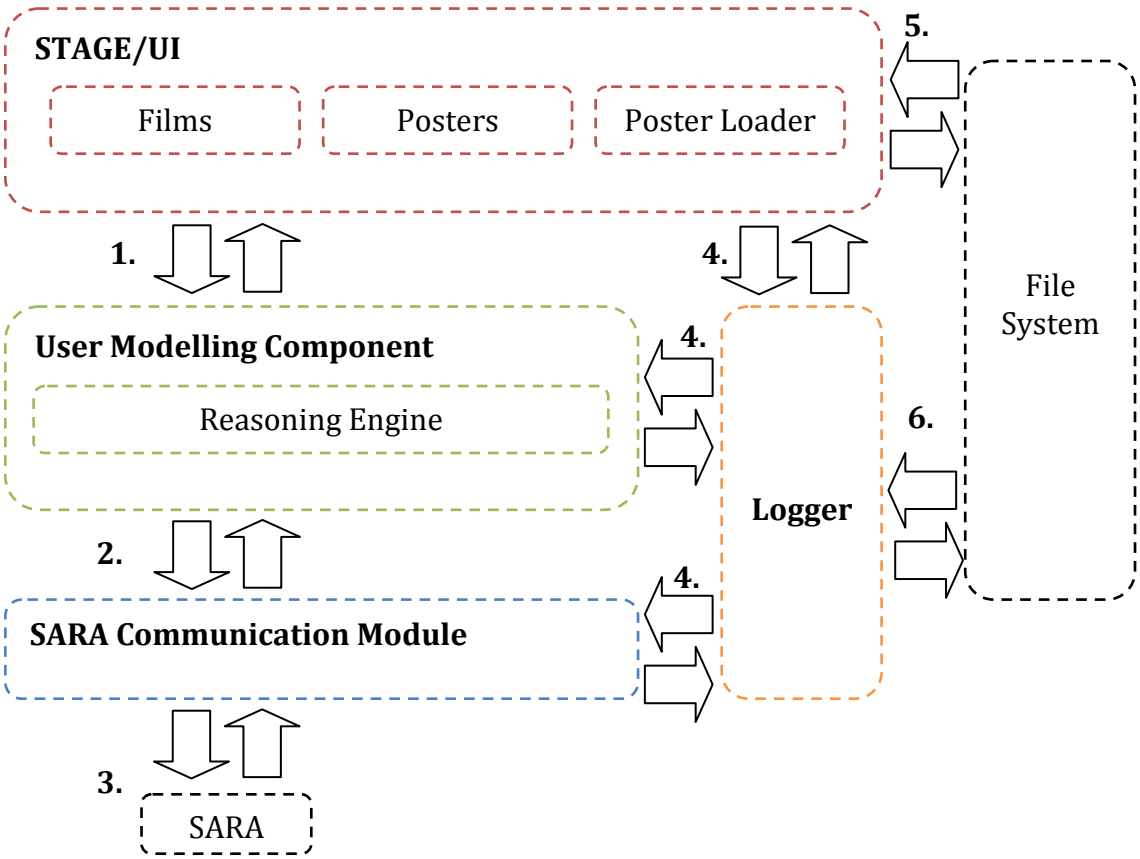


Figure 33: System Architecture and Communication Patterns

4.4.3 Implementation of FDEC

The Film Domain Exploration Client has been developed using Adobe Flex Builder 3. As such it uses ActionScript 3 (AS3) which is a scripting language based on ECMAScript [**78**]. The Adobe Flash CS4 IDE has been used to design numerous visual instances of the constituent parts of the posters which are then loaded into the application when needed. Colour coding and testing of numerous filter properties (glowing, shadow, transparency) to achieve an organic flow and visual appeal has been also done within the Flash IDE.

Originally, the Film Domain Exploration Client consisted of 3 steps. In step 1 a user defined his preference film. Step 2 displayed randomly selected film to be rated in favour or against. Step 2 was removed since it didn't prove to be very helpful and the same results could be achieved within the main part of the application. Figure 34 shows the main stage from a distance.

Figure 34: Screenshot of FDEC showing the main stage with rated items

Figure 34 further shows one central interaction paradigm: Film posters that have been rate differently appear glowing in different colours.

The rating is straight forward and allows only three different stages: A user can either like, dislike or mark a poster as neutral. By doing this the User Modelling component

is able to learn from these interactions and build and refine a model of the user. How this is technically achieved is discussed next in more detail.

4.4.4 User Modelling and Use Case

The core of the implementation is the User Modelling component. Its job is to collect and evaluate user interactions with the system. The very first interaction it registers is when a user sets his initial focus film.

This film specifies what types of film posters are going to populate the stage. Keeping in mind that there are a total of 1,800 films in the database, we integrated some uncertainty when acquiring films. This helps to prevent repetition and assists with the objective of showing new films a user may be interested in.

Otherwise the reasoning behind is fairly simple. At the start there will be 20 poster images on stage. We are using four different semantic attributes. Hence, the number of posters with regard to any semantic attribute is 5 as shown in lines 10-14 of Figure 35.

```
*****************************************************************************************
** New Session: Wed Sep 2 02:31:28 GMT+0100 2009
*****************************************************************************************
[1] 00:00:48: Initial Focus Movie: memento
[2] 00:00:48: Liked films (total, multiple rounds): 0
[3] 00:00:48: Neutral films (total, multiple rounds): 0
[4] 00:00:48: Disliked films (total, multiple rounds): 0
[5] 00:00:48: Total rated films: 0
[6] 00:00:48: PopularGrossingWorld(liked, neutral, disliked): 0,0,0
[7] 00:00:48: AwardWinner(liked, neutral, disliked): 0,0,0
[8] 00:00:48: PopularGrossingWorldBudgetRatio(liked, neutral, disliked): 0,0,0
[9] 00:00:48: PopularRating(liked, neutral, disliked): 0,0,0
[10] 00:00:48: Current Sem Atts (weighting): PopularGrossingWorld: 5
[11] 00:00:48: Current Sem Atts (weighting): AwardWinner: 5
[12] 00:00:48: Current Sem Atts (weighting): PopularGrossingWorldBudgetRatio: 5
[13] 00:00:48: Current Sem Atts (weighting): PopularRating: 5
[14] 00:00:48: Sem Att PopularGrossingWorld (SARA query parameters): 35699355.21212,43632545.211212
[15] 00:00:48: Sem Att AwardWinner (SARA query params):
false,false,false,false,false,false,false,false,false,false,false
[16] 00:00:48: Sem Att PopularGrossingWorldBudgetRatio (SARA query parameters):
6.3112424,7.711241210000001
[17] 00:00:48: Sem Att PopularRating (SARA query parameters): 8.181234324,9.04124324
[...]
```

Figure 35: Session log file

It is important to note, that this number can vary in some rare occurrences, especially when SARA is not able to acquire and return the number of films it is asked for. This can happen for example when there are not enough films based on certain parameters and when these parameters are too narrow.

A query to SARA includes the following steps:

1. Collect information about focus film.
2. Check User Model.
3. Query SARA for films similar to the characteristics of the current focus film, but define the number of expected return items based on the User Model and the current weighting of semantic attributes correlated with it.
4. Receive result and update screen objects.

The query parameters for SARA with regard to the current focus film can look like represented in lines 14 - 17 of Figure 35.

The values in lines 14, 15 and 17 define the lower and upper limits when SARA searches for similar films. These boundaries are set in the region of plus/minus 10-20 percent.

A query formed within Flex Builder 3 is shown in Figure 36. It gives an idea of how semantic attributes can be personalised on the fly by adjusting their parameters.

The semantic attributes are balanced as follows: In order to change the weightings of semantic attributes, the user has to explicitly rate films. The main objective is simply to "*find films you like*". By doing any ratings, a user implicitly tells the application which semantic attributes he prefers and which not, because every poster that is shown on stage has a semantic attribute attached to it (and this is the reason why it has been returned by SARA originally).

This happens totally transparent to the user who only interacts with the poster. The User Modelling component collects these preferences. It further distinguishes between the current round starts by setting a new focus film, and the entire session. This allows the system to learn, shift and update the user model and strongly bend toward one direction at one time but toward a different one at some other stage. Having some sort of a history functionality in place as well as by using some Bayesian

algorithms [**97**] it is possible to keep a "short tail". This allows the user to escape extensive temporal shifting. A single semantic attribute is never removed completely. Based on the total number of films that should be presented on stage at any single time (which is in the area of 30), an equal share among the (four) semantic attributes is calculated.

```
public function getPopularGrossingWorldBudgetRatio(n:int = 30, l:Number = 0.1, u:Number = 900000000):void{

        var singleSemAtt:ArrayCollection = new ArrayCollection();
        //create single semantic attribute

        singleSemAtt.addItem("PopularGrossingWorldBudgetRatio");
        //should match semAtt name specified in authoring tool

        singleSemAtt.addItem("Default");
        //should match the parameter name specified in the semAtt authoring tool. If instance or aggregate this
        is "Default"

        singleSemAtt.addItem(true);
        //is this sematt currently personalised or not

        singleSemAtt.addItem("Number");
        //whatever the variable datatype associated with the SemAtt is. currently only a single datatype can be
        associated with a semantic attribute but can easily be changed if necessary

        singleSemAtt.addItem("movie");
        //will always be movie

        singleSemAtt.addItem(n);
        //number of wanted results back

        chosenSemAtts.addItem(singleSemAtt)  //add instance to master array

        var singleSemAttArgs:ArrayCollection = new ArrayCollection();
        //create single semantic attribute arguments array. selected tags that users want, will go in here.

        singleSemAttArgs.addItem(Number(l+0.00112424));
        //lower bound; add any tiny float number to fix bug with built-in number conversion in AS3

        singleSemAttArgs.addItem(Number(u+0.001124121));
        //upper bound

        log.write("Sem Att PopularGrossingWorldBudgetRatio (SARA query parameters): " + singleSemAttArgs)

        //write to log file

        parameterArgs.addItem(singleSemAttArgs) //add it to master

}
```

Figure 36: Single Semantic Attribute query in ActionScript3

The approach that I explained above in combination with continuous user exploration was used to build a comprehensive model. Subjective expert knowledge

is represented through semantic attributes. The log files allow seeing and evaluating at what point a certain user agreed with what type of attached attributes. This further relates to pre encoded expertise because shown films can be related back.

Figure 37 shows the effect of rating posters by using selected passages from a log file that is created with every session. It clarifies what is going on when a user rates a certain semantic attributes favourably. You can see how the number of films related to it changes with relation to the overall ratings, session ratings, number of semantic attributes and available films.

When there are a number of ratings per round, it quickly becomes unclear what is going on. It is important to note, that I deliberately chose a simple session log to avoid confusion with the reader. Additionally I removed irrelevant lines and highlighted important sections to improve legibility. For complete and more complex examples I refer to Appendix E.

****
** New Session: Sat Sep 5 12:22:55 GMT+0100 2009
****
[1] 00:00:18: Initial Focus Movie: The Godfather
[...]
[10] 00:00:18: Current Sem Atts (weighting): **PopularRating: 5**
[11] 00:00:18: Current Sem Atts (weighting): **PopularGrossingWorld: 5**
[12] 00:00:18: Current Sem Atts (weighting): **AwardWinner: 5**
[13] 00:00:18: Current Sem Atts (weighting): **PopularGrossingWorldBudgetRatio: 5**
[...]
**** New focus movie **********
[18] 00:00:31: **Rating: 1 | PopularRating** | The Silence Of The Lambs
[19] 00:00:32: **Rating: 1 | PopularRating** | Indiana Jones And The Last Crusade
[20] 00:00:33: **Rating: 1 | PopularRating** | Die Hard
[21] 00:00:35: **Rating: 1 | | The Godfather**
[22] 00:00:38: Current Focus Movie: The Godfather ()
[...]
[31] 00:00:39: Current Sem Atts (weighting): **PopularRating: 12**
[32] 00:00:39: Current Sem Atts (weighting): **PopularGrossingWorld: 3**
[33] 00:00:39: Current Sem Atts (weighting): **AwardWinner: 3**
[34] 00:00:39: Current Sem Atts (weighting): **PopularGrossingWorldBudgetRatio: 3**
[...]
**** New focus movie **********
[39] 00:01:05: **Rating: 1 | PopularRating** | Fight Club
[40] 00:01:07: **Rating: 1 | PopularRating** | Sin City
[41] 00:01:07: **Rating: 1 | PopularRating** | Up
[42] 00:01:09: **Rating: 1 | PopularRating** | No Country For Old Men
[43] 00:01:11: Current Focus Movie: Up (PopularRating)
[...]
[52] 00:01:14: Current Sem Atts (weighting): **PopularRating: 16**
[53] 00:01:14: Current Sem Atts (weighting): **PopularGrossingWorld: 2**
[54] 00:01:14: Current Sem Atts (weighting): **AwardWinner: 2**
[55] 00:01:14: Current Sem Atts (weighting): **PopularGrossingWorldBudgetRatio: 2**
[...]

**** New focus movie **
[60] 00:01:38: **Rating: 1 | PopularRating** | Star Trek
[61] 00:01:40: **Rating: 1 | PopularRating** | The Prestige
[62] 00:01:49: **Rating: 1 | PopularRating** | Alien
[63] 00:01:53: **Rating: 1 | PopularRating** | Slumdog Millionaire
[...]
[73] 00:01:59: Current Sem Atts (weighting): **PopularRating: 19**
[74] 00:01:59: Current Sem Atts (weighting): **PopularGrossingWorld: 2**
[75] 00:01:59: Current Sem Atts (weighting): **AwardWinner: 2**
[76] 00:01:59: Current Sem Atts (weighting): **PopularGrossingWorldBudgetRatio: 2**
[...]
**** New focus movie **
[81] 00:02:37: **Rating: 1 | PopularGrossingWorldBudgetRatio** | Beverly Hills Cop
[82] 00:02:42: **Rating: 1 | PopularGrossingWorldBudgetRatio** | Top Gun
[83] 00:02:51: **Rating: 1 | AwardWinner** | The Exorcist
[84] 00:02:54: **Rating: -1 | PopularRating** | Murderball
[85] 00:03:08: **Rating: -1 | PopularRating** | The Truman Show
[86] 00:03:13: **Rating: -1 | PopularRating** | Mary Poppins
[87] 00:03:29: **Rating: 1 | AwardWinner** | Forrest Gump
[88] 00:03:41: Current Focus Movie: Donnie Darko (PopularRating)
[...]
[97] 00:03:45: Current Sem Atts (weighting): **PopularRating: 15**
[98] 00:03:45: Current Sem Atts (weighting): **PopularGrossingWorld: 2**
[99] 00:03:45: Current Sem Atts (weighting): **AwardWinner: 5**
[100] 00:03:45: Current Sem Atts (weighting): **PopularGrossingWorldBudgetRatio: 5**
[...]
**** New focus movie **
[105] 00:05:10: **Rating: 1 | PopularGrossingWorld** | Glitter
[106] 00:05:12: **Rating: 1 | PopularGrossingWorld** | The Good Thief
[107] 00:05:16: **Rating: -1 | AwardWinner** | Where The Truth Lies
[108] 00:05:18: **Rating: -1 | AwardWinner** | Red Dragon
[109] 00:05:21: **Rating: -1 | AwardWinner** | The Legend Of Bagger Vance
[110] 00:05:22: **Rating: -1 | AwardWinner** | Igby Goes Down
[111] 00:05:31: Current Focus Movie: The Good Thief (PopularGrossingWorld)
[...]
[120] 00:05:34: Current Sem Atts (weighting): **PopularRating: 15**
[121] 00:05:34: Current Sem Atts (weighting): **PopularGrossingWorld: 6**
[122] 00:05:34: Current Sem Atts (weighting): **AwardWinner: 3**
[123] 00:05:34: Current Sem Atts (weighting): **PopularGrossingWorldBudgetRatio: 6**
[...]
**** New focus movie **
[128] 00:07:51: **Rating: -1 | PopularRating** | The New World
[129] 00:07:52: **Rating: -1 | PopularRating** | The American President
[130] 00:07:53: **Rating: -1 | PopularRating** | Session 9
[131] 00:08:04: **Rating: 1 | PopularGrossingWorld** | Unknown
[132] 00:08:07: **Rating: 1 | PopularGrossingWorld** | Cradle Will Rock
[133] 00:08:08: **Rating: 1 | PopularGrossingWorld** | Supercross
[134] 00:08:10: Current Focus Movie: Supercross (PopularGrossingWorld)
[...]
[143] 00:08:13: Current Sem Atts (weighting): **PopularRating: 10**
[144] 00:08:13: Current Sem Atts (weighting): **PopularGrossingWorld: 12**
[145] 00:08:13: Current Sem Atts (weighting): **AwardWinner: 5**
[146] 00:08:13: Current Sem Atts (weighting): **PopularGrossingWorldBudgetRatio: 6**
[...]

Figure 37: Log file showing the user modelling based on reconciliation

4.5 Implementation Issues and Discussion

The process of gathering all the necessary data was way more extensive than expected. One would consider that data mining in this area due to an abundance of different sources would be unproblematic. But it was important to have enough variation from disparate supply sources. This resulted in having to deal with numerous different systems and data formats which consumed a big part of the time available for the whole project.

I am moving over to talk about the software implementation: At first there was some uncertainty about which IDE and technology should be used. It was undecided whether it would be Adobe Flash CS4 or Flex Builder 3. Both of them use AS3 as their coding language but exhibit some major differences when it comes to setting up a project. Eventually, it was decided to go with Flex in order to utilise BlazeDS and connect to SARA which runs within Tomcat.

To make it simple for distribution, the FDEC is distributed as an Adobe Air Application. This allows it to be installed and run on numerous platforms and on any machine where the appropriate backend architecture is in place. Changing the client to be run from within a web browser can be also easily accomplished.

The system is able to handle more than just the four semantic attributes that are currently implemented. If there are more than four semantic attributes, the application would break the number evenly apart. The axis used right now is “popularity”. It does not accommodate another axis of semantic attributes.

It is challenging to come up with further types of semantic attributes than just the ones discussed earlier in this chapter. More subjective user generated meta data is difficult to parse and access programmatically.

Specific film characteristic like colour coding, colour temperature, number of cuts per second or different camera angles may exhibit a base to support the creation of very interesting semantic attributes. Once again, these information may be quite difficult to gather. This could be rather difficult with domains dissimilar to film or media, too.

4.6 Conclusion

The implementation chapter has shown what steps have been involved in creating the Film Domain Exploration Client. To a great extent the process includes data acquisition, pre processing and the creation of semantic attributes. The chapter also focuses strongly on the technical implementation and describes what technologies have been used and for what reason. Further it assesses the architecture and the numerous building blocks within the application. It should be comprehensible how distinct high level branches like “exploration”, “personalised expert perspectives”, and “user modelling” are glued together successfully. Problems and encountered challenges are also pointed out and solutions are suggested.

5 Evaluation

5.1 Overview

This chapter assesses the project. In user trials The Film Domain Exploration tool is evaluated along numerous parameters. These parameters comprise usability with regard to the visual user interface component and its metaphors as well as user perception. The general value of the tool is another issue. Moreover, it is evaluated whether and to what extent research goals and objectives as projected in chapter 1 have been reached. All in all it was asked whether implicit and explicit user modelling techniques can be appropriately applied to balance different subjective expert perspectives to aid visual exploration of media resources. Can this approach improve the exploration of large volumes of information where sufficient metadata is obtainable?

There were two rounds of user trials with different objectives. The first round was conducted over a period of 2 days and involved seven participants. Round number 2 was held after the first session had been evaluated and interesting results had been collected. The second test group comprised a deliberately selected set of 3 people who had given motivating comments, insights or assumptions with regard to the user interface and to the underlying processes. All of them tried at least at once to intentionally rate and focus on films different to their taste to trick the system and get a deeper insight of what was going on. Another reason to be chosen to participate in a second round was if a user exposed negative thoughts or exceptionally critical comments.

5.2 Evaluation 1

5.2.1 Design

To be able to gain detailed information about a user's interaction, two recording systems have been used. For one, the application writes a comprehensive log file which includes all the important session and user modelling data as computed within the Film Domain Exploration Client. In addition to that, video capturing software was

used to record mouse and click paths. In combination with the log files it was possible to synchronise timings and draw further conclusions. Moreover, all participants have been encouraged to express their thoughts openly to collect extra information that would otherwise not be documented. Afterwards a questionnaire (Appendix F) was provided which invited the volunteers to answer 49 questions spanning quantitative and qualitative questions about "films in general", software client and UI perception as well as universal statements about assumptions, feelings and comments.

Although I implemented the feature to hit the TAB key to pick a random film in stage 1, I deliberately chose not to tell any participants. Instead I preferred that they would come up with a film on their own. That way I wanted to force them to get emotionally started with thinking about films as soon as possible.

5.2.2 Execution and Evaluation

After a brief introduction which was supported by using a couple of power point presentation slides to provide identical information to everybody upfront, the participants were asked to interact with the client application.

The first round of trials consisted of seven volunteers. Five of these were male, and two were female. All of them were of age 24 to 30. Apart from one user who considered himself "somewhat" comfortable with computers and another one indicating moderate familiarity with computers, all participants have been "very much" into computers.

"Find 25 to 30 films you like through exploring!" was the main and only task to complete. It was knowingly kept fairly simple to eliminate any form of pressure or suspicion of a deeper objective. The user should be as ignorant as possible.

A trial session took approximately 40 - 55 minutes. This included introduction, using the application and filling out the questionnaire. The interaction with the tool took between 10 and 30 minutes; the majority 20 minutes. The reason for this was that it was assumed that a participant's immersion with the system indirectly supports its success. Therefore it could be carefully stated that the original task of finding a specific number of films receded towards a continuous exploration of the poster images that have been presented. This can be backed by the fact that even after

participants have been asked for the number of films that they had rated so far and whether they reached their goal, no one was able to give a definite answer. Since I haven't asked them straight forwardly to finish the session, apart from one participant, everyone kept on being strongly focused and continued using the tool.

Next, I am going to evaluate the answers given with regard to the film domain in general. Everybody agreed to a certain degree that he was interested in films. All but two confirmed to know a lot about them. The remaining two stated that their knowledge depended on the film genre. This could be backed up by the highly subjective fact that most of them agreed they would be watching more films now that they did 10 years ago and as such agreed. Only one volunteer disagreed and said he was watching not that many films lately but TV series instead. But this exhibits a strong association to films, too.

Half of the test group emphasised that they would not get information about films on a regular basis. However everyone confirmed that the first place he came to hear about new films was online. Half of the group also used "friends" as an important source of information. Only one participant additionally mentioned magazines and another one added TV. With regard to how consistently they used online services or websites to gather information about films all but one agreed to do this often or very often. This was the one who stated his only "somewhat" familiarity with computers earlier which could be a high level explanation for answering the question this way.

Film domain

After that the questionnaire covered available film recommendation sites on the internet. The results can be summarised as follows. IMDb seems to be everyone's number one choice. Rotten Tomatoes was marked by more than half of the group when it came to regular use. All the other listed websites like Freebase, MovieLens, What to Rent, Criticker, Clerkdogs or Jinni are ignored at all. When explicitly asked to put down any further resources, participants mentioned the following sites (starting with the ones referenced more often): Netflix, ScreenClick, Wikipedia.

This is interesting since they appear to care a lot about films, but concentrate on pretty much one or two sources only. This can be evaluated in two ways: For one, it might be difficult to change the people's habit of how they usually get information, on

the other side there is potential of my application since its approach is rather different.

When looking for film related information, more than half of the participants used sources that are described above and primarily to get general information.

The personal perception of what they like and what they do not was fairly divergent. With regard to "blockbuster films" more than half of the group agreed to one extent. The rest nearly seemed to be offended by the question, disagreed and emphasised on their "indie taste". A similar pattern can be applied to "expensive films", "award winning films" and "successful films". Participants usually slightly agreed or disagreed but were happy to list what their own categories would be. This was quite different and no real focus could be found. People like "sci-fi", "thrillers", "action", "films that are based on true stories", "(Spanish) comedies", "British films with dark humour", "weird arthouse", "films with Non-Hollywood themes", "strange and underground films" as well as "entertaining" and "thought-provoking" ones. Taking this into consideration a film recommendation or exploration tool has to accommodate a variety of different fields of interest and people did not tend to agree to the predefined genres which have been derived from the semantic attributes used in the application.

User Interface

The following paragraphs contain results that have been drawn with regard to the actual implementation. In first place, user interface issues are evaluated.

All participants strongly agreed that it was easy to select an initial film. Apart from one who strongly disagreed, everyone knew that he was able to change the initial film before continuing. All volunteers agreed that it was simple to rate films. However, everyone pointed out not to be content with the rating metaphor and preferred that clicking on the upper part of a poster image would rate it positively and vice versa. The 3D metaphor of pushing and pulling wasn't recognised. Another valuable comment was that the arrows that appear on top of a poster and indicate how it would be rated should be changed towards more intuitive symbols that represent "plus" or "minus". Every participant still highlighted that although this was confusing in the beginning and should be definitely changed the current rating scheme was still

easily understood and worked well. Everyone strongly agreed that setting a new focus film was simple to do and zooming very intuitive. Furthermore, panning the stage vertically and horizontally to reach different areas of the stage was strongly agreed to be easy and straight forward.

With regard to the question whether it was easy to get lost, interesting results occurred. Although it was part of the UI section, the statement apparently was a little bit too ambiguous and misleading. It was meant to ask whether it was easy to reach a point where the user did not know how to continue. This question was included to evaluate whether the interface would get too complicated at any stage. After some clarifications, all volunteers disagreed or disagreed strongly. However, more than half of the test group interpreted the question with regard to the level of immersion. Comments about whether it was easy to get lost were "Sure, that is the idea!" or "...but in a good way – exploring!". One participant noted that "IMDb is not good at this." Another excellent suggestion at this stage was that it would be helpful to include an arrow at the edge of the screen towards the centre.

The visual representation of film posters and title was considered useful and helped identifying the posters. One volunteer disliked the white border around the poster images and wanted the space between the posters in general to be smaller.

Questions about exploration, discovery and the interface in common were as follows: The majority agreed (some of them even strongly agreed) that it was easy to find the films that have been rated in the session and that the user interface made it rather easy to explore films. One volunteer complained that he was "trapped to [his] initial" film.

Everyone agreed that the interface was fun to use. One participant commented on the "spiral effect" which he really liked, because it helped identifying the focus film and the relations on screen.

The main suggestions with regard to what has to be improved or changed, that had been stated at least twice, are as follows: Participants wanted to see a button to quickly browse the list of liked and disliked films. Zooming out to access rated films should be changed. The ability to ignore certain films should be implemented. A

search bar which allows for text input selecting any focus film should be available throughout the whole application.

Rating is confusing and should be changed. The push/pull metaphor is to be changed towards a more intuitive plus/minus metaphor. Navigational features should further allow the use of the keyboard.

Additional questions and general statements

Below I will detail about the last part of the evaluation in context with general impressions of the user interface, the application and usability issues. Observations and further user comments will be also discussed.

With regard to whether it was difficult to come up with an opening film in step 1, different answers had been given. Half of the volunteers agreed, but the other half strongly disagreed. Someone mentioned that he wasn't able to find a certain film ("Lock, Stock and Two Smoking Barrels") since it wasn't included in the system and as a result had to pick another one. An improvement of the auto complete functionality was suggested, because the current implementation should handle films starting with "the" which are currently a reason for confusion, "but afterwards it was all cool". At the moment for example "The Usual Suspects" are different to "Usual Suspects". One participant thought it would be nice to have random films fading in and out in the beginning to get some first ideas and guidance. Apart from that everybody agreed that the system worked well.

All participants (strongly) agreed to have been finding films they liked. Furthermore, nearly everyone agreed to have found films he expected. The two other ones disagreed and commented to have been "surprised" most of the time. One of them stated in addition that being surprised was "not a bad thing" essentially. He wanted to express that he liked the fact to see films he wasn't intending to be looking for. When being asked for opinions about the system's reasoning and predictability, a wide range of different responses came back spanning from "I strongly disagree" to "I strongly agree". I quote: "I didn't find any relations between the film in the middle and the rest", "What does 'X' have to do with 'Y' and why should I like 'Y'?" or "Refocusing brought a new selection of related movies I like, with irrelevant movies

at the outer edge!". Another interesting response was that the system was predictable, but it wasn't clear what was going on "under the hood".

To evaluate the level of distraction the participants were asked to put down how many times they thought they refocused on a new film. A few didn't recall at all. The majority was able to give correct answers which were usually among 6 - 8. Towards a general perception of the Film Domain Exploration Client, people found that it wasn't a waste of time interacting with it - even if they didn't figure out how the presented results are related. They were happy with the films that had been shown to them. One volunteer stated that he would have loved to spend even more time with the application.

Questions that aimed to additionally check the people's experience are evaluated next. The majority emphasised that it had been shown films it hasn't been thinking about for a long time. Furthermore, the effect of being presented with these films made the participants want to watch one of the films they found right away if they had time.

In the last few questions I once more explicitly asked the volunteers what they though, the application was good for, or what it could be used for. A number of annotations perfectly highlight the second-level objectives of the implementation. Comments were: "Film exploration and getting a map of everything you have seen.", "Suggesting a movie and then finding a movie to watch", "Finding recipes! The system is based on some logic which is bringing up unexpected results. It's like to be surprised continuously!", "Wish list!", "Finding new films to watch, especially ones you haven't heard of and haven't seen!", "Ability to find resembling movies that either you haven't seen or you missed!" or "Good to remember films that you didn't think of at the moment. It can be used as a favourites list!"...

Further issues

The test group provided a number of additional comments that could improve the overall user experience of the application a lot more: Relations between shown films and new films after resetting focus are difficult to understand and it was wished to be able to remove films that have been rated already and also to have more clarity with regard to how to access liked and disliked films.

"Link films to external sources like IMDb or rental stores would make it easy to select films you may like", "Add additional information to films to assist with decision making" and "Add comments to any movie" was mentioned a few times. Also "Browse movies by genre and other criteria" was high on the wish list. In general, people wanted to have more assistance with the initial selection of films.

Participants liked the visual representation and user interface. Concluding remarks covered quotes like:

- "I like the exploration!"
- "Could play with it for hours!"
- "System returned movies I haven't seen in a long time."
- "I like the variety of shown poster images."
- "Interesting for recollecting your movie data"
- "Easy to build catalogue of liked/disliked films. IMDb does not do this."
- "I'm still curious about the reasoning behind!"
- "Nice application! Could use it!"
- "Cool idea!"
- "Put into IMDb now!"

5.2.3 Summary

During the course of the trial more than half of the group suspected some deeper reasoning behind, but wasn't entirely able to really guess what it was. Most participants (apart from one who couldn't grasp what is going on at all) correctly assumed the reasoning behind takes some sort of user ratings and connected meta data into account to calculate what to show next. However, suggestions primarily included genre, actors or directors.

The results that have been gathered in this first round exhibit two different general characteristics: Apart from user interface and interaction issues which created some confusion, all volunteers were able to gain sufficient immersion and enjoyed using the tool exploring films.

It has to be noted that the results from the usability section could have been different if there was no introduction or personal assistance at all. The application still

requires changes towards better usability to be understood by less experienced people with regard to HCI.

On the other side however, no one was actually able to figure out what was going on behind and how the films on stage were actually related to each other. The same answer was given with regard to the focus film and to the earlier rated films. There was no clear consistency to be seen. The findings in this first session of user trials served as a perfect starting point for a second round which will be discussed next.

5.3 Evaluation 2

5.3.1 Design

The second session was done after the first was evaluated. I selected 3 participants from round 1 and told them what the system is actually doing and how the components are tied together. A single session took about 10 to 15 minutes.

The key part of these was to further evaluate the users' perception of the system and what it essentially modelled and to collect subsequent reflective responses. Furthermore, it was crucial to see whether a user was able to see the correlations and his opinions about these when confronted with modelling results from the log files.

5.3.2 Execution and Evaluation

I started with replaying short but distinct passages (4-6 minutes) of the videos I had taken in round 1 and asked the second round participants about their intentions. All answers expressed more or less that the users didn't know what they were looking for and simply followed the given task and their urge of exploration.

Afterwards I presented and explained the log files that had been created during the first sitting. To every volunteer I pointed out a number of iterations in their user modelling process. Then I explained in more detail how to properly interpret the logs.

Initially, the topic appeared to be rather difficult to understand for all three participants. The novel approach of combining expert knowledge encoded in

semantic attributes along with user modelling techniques seemed to be rather confusing and understanding the idea behind semantic attributes, how they can be combined and linked together along specific axes, was a challenge to communicate. Especially to have the users see why semantic attributes are different to simple tags was difficult.

I continued to step through the log and explaining the modelling process that exhibited the acquired results. At certain stages I confronted the participants for example that the system thought they were into popular films, and at another time I highlighted that they were considered to be looking for award winning films for example. With the films, ratings and comprehensible results in front of them, I asked whether they would agree to a certain modelling outcome. All participants negated. No one was actually able to see himself looking for any of the four semantic attributes that have been used in this project. As outlined in 3.7 these attributes were "popularity rating" as rated on IMDb, "grossing worldwide" based on box office results, "award winners" as of Academy, BAFTA, Cannes, Berlin, MTV, and the "ratio between worldwide grossing and budget".

In the end I wanted to know if they would change their opinions about the application with regard to what they stated during the first round of trials. Everyone said no right away. Two volunteers emphasised on the fact that they were happy to know about the internal reasoning although their personal presumption was totally different. The newly gained knowledge however didn't change the fact that they liked the non-invasive character of the exploration and that they were able to find films.

5.3.3 Summary

The volunteers were picked because the log files indicated different levels of clarity with respect to balancing and modelling. The results from the second trial can be interpreted as follows: There is no apparent indication that the user model represents what the users had perceived. This can be explained by the fact that the users were unfamiliar with the reasoning and modelling approach. Moreover, they assumed that the modelling was based on different criteria and/or just based on simple ones like similar actors or producers. They would not agree with the semantic attributes that have been used within this application. A few times, participants

pointed out that they experienced some sort of change or guidance regarding the films that the applications showed to them. But once again no obvious connection could be expressed.

5.4 Conclusion

There are a number of different conclusions that can be drawn with respect to different components of the project. Having two degrees of subjectivity as what users like and what experts consider to be mapped, makes it difficult to evaluate the degree of success.

There have been a couple of suggestions for improvement, but it can be said, that all participants liked the user interface and navigation in general. The group members agreed that they were able to complete the task of finding films they like and felt a strong immersion while doing so.

Confronted with the technical details of the modelling and balancing approach the participants stated to be unaware of the system's reasoning. But it was definitely seen that the implicit modelling supported by subjective semantic attributes worked from a technical perspective.

One participant remarked that it would not be of great importance to him to be ignorant of internal logics and balanced assistance as long as the results are plausible. However, it was mentioned also that it would be helpful if modelling results would be conveyed more explicitly by the application.

Using different semantic attributes or more variety exhibit some potential. It would make it easier for the test group to understand the user modelling processes since their preferences would be reflected in a more analogous way.

6 Conclusion and Future Work

6.1 Overview

This chapter summarises and assesses how this work has achieved the research statement, objectives and goals that have been defined in chapter 1. Specifically it discusses the results achieved and their potential. It also proposes areas of future investigation stemming from this work.

6.2 Motivation and Objectives

The motivation for this project was to access and handle large amounts of data in a semantically meaningful, and for the user non-overwhelming, way. In detail it was examined whether implicit and explicit user modelling techniques could be applied to balance different subjective expert perspectives in a highly graphical and pervasive exploration environment. For this reason the film domain was chosen. The dissertation evaluated how this approach improved the exploration of large information volumes, where sufficient metadata is available.

A number of objectives and goals were defined to guide through this project. In chapter 2 the relevant State of the Art was analysed. This included Knowledge Discovery and Data Mining (KDDM) techniques, Flexible Querying and Dataspaces which lead to a close investigation of the Semantic Attribute Reconciliation Architecture (SARA) and its specific process model. It was highlighted that data understanding is the central component and how this lead to the invention of semantic attributes. Semantic attributes represent subjective characteristics of a domain encoded by experts. These properties can be personalised to an end user's preferences and context. They enable users to make semantically meaningful and expertly enriched queries across heterogeneous information sources. By balancing and throttling the influence of these attributes the application was able to react to user choices and towards exploration.

Implicit and explicit user modelling techniques and how these could be balanced appropriately by utilising semantic attributes was evaluated next, which informed how the Film Domain Exploration Client modelled the user's interests based on

expert's guidance expressed in semantic attributes. Special attention was also paid to current state of the art film recommendation systems as well as compelling visualisation techniques. It was shown what features they lack, how these influenced the design of the Film Domain Exploration Client and why this approach can be considered to be novel.

As part of another important step, heterogeneous information sources were retrieved and organised throughout a lengthy process. This was necessary in order to build a base set of data and to be able to define semantic attributes. This information was used later on during the personalisation procedure for balancing the semantic attributes within the user application. In the evaluation, challenges when modelling a user's intent were surveyed. It was furthermore assessed how these interactions with the system were reconciled to update the user model. Moreover, implicit and explicit feedback from users was collected.

6.3 Contribution

The project offers an alternative to collaborative filtering and investigated the challenges when modelling a user's intent. By interactively exploring a knowledge space using a highly visual interface, the user was not burdened with unnecessary modelling and information seeking concerns in order to maintain flow and immersion. The balance between explicit and implicit modelling techniques was maintained to ensure the user had a contiguous experience. The personalised nature of the interactions and the user's subsequent discoveries presented a need for detailed, yet quickly changing user model.

- Since it invites exploration, a compelling and functional user interface was significant for this type of modelling approach.
- For the application to being useful, data was collected and collated upfront. This consumed and great deal of time and resources.
- Once the step above had been accomplished, inventing and defining new semantic attributes was fairly effortless.
- It was shown that semantic attributes gave users different sets of criteria to explore the film domain. The fact that these attributes overlap, necessitates careful balancing.

- It is imperative to create a greater number of semantic attributes than this work was able to produce. A bigger variety and amount of retrieved meta data would enable this. Only then is it possible to cover a multiplicity of different axes.
- The implicit modelling aspect triggered by explicit user actions worked and successfully rebalanced the user model.
- A correlation of the focus film to other ones and films that have been displayed or selected during earlier iterations within one session was not apparent to users. This has to be addressed on the UI level.
- When applying expert perspectives to support user exploration, an efficient form of scrutability is required [**39**]. This is even more fundamental when there are several axes of semantic attributes involved.
- Although users were not aware of the assistance given by the application they were able to navigate and find films they liked.
- Awareness of the underlying reasoning as such is not a key aspect for users. They were able to successfully navigate the application and find films even without realising that they were assisted by the system or did not care about it.

6.4 Future Work

Evaluation and user comments brought up numerous suggestions about improvement of the application. Most of these issues were related to UI or navigation. Still, the interface worked and allowed successful exploration.

It is a crucial part to cover implicit and explicit user modelling techniques. Furthermore incorporating and balancing expert perspectives to allow guidance, offers a lot of opportunities. It should become apparent that the discussed approach can be applied to a number of other domains as long as sufficient heterogeneous meta data can be acquired. This dissertation elaborates that there is great potential in pursuing these in more detail.

Future work in this field will provide advanced ways of exposing the underlying logic in a more explicit way and consequently allow users to get imminent feedback about the systems' modelling processes.

Another very important aspect of gaining superior modelling results, definitely facilitates improvements when it comes to the number of axes and range of different semantic attributes.

It is obvious that a lot of further research can be done in any of the areas stated above. Current systems do not offer the approach proposed in this dissertation to improve exploration of large volumes of information. It has been revealed that there is great potential when it comes to personalisation as well as to effectively applying implicit and explicit user modelling techniques to balance various subjective expert perspectives and to aid visual exploration of (not just) media resources.

Bibliography

[1] Alfred Kobsa, "User Modeling: Recent Work, Prospects and Hazards," in *Adaptive user interfaces: Principles and practice*, T. Kühme and U. Malinowski Hufschmidt, Ed. Amsterdam, North-Holland, 1993, pp. 111-128.

[2] Michael S. Lew, Nicu Sebe, Chabane Djeraba, and Ramesh Jain, "Content-based multimedia information retrieval: State of the art and challenges," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, vol. 2, no. 1, pp. 1-19, 2006.

[3] (2009, August) The Internet Movie Database. [Online]. http://www.imdb.com/

[4] (2009, August) MovieLens. [Online]. http://www.movielens.org

[5] (2009, August) Rotten Tomatoes. [Online]. http://www.rottentomatoes.com/

[6] Richard E. Mayer and Roxana Moreno, "Nine Ways to Reduce Cognitive Load in Multimedia Learning," *Educational Psychologist*, vol. 38, no. 1, pp. 43 - 52, March 2003.

[7] Lukasz A. Kurgan and Petr Musilek, "A survey of Knowledge Discovery and Data Mining process models," *The Knowledge Engineering Review*, vol. 21, no. 1, pp. 1-24, 2006.

[8] M. Goebel and L. Gruenwald, "A survey of data mining and knowledge discovery software tools," *SIGKDD Explor. Newsletter vol.1*, pp. 20-33, 1999.

[9] William B. Rouse, "Need to know-information, knowledge, and decision making," *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, vol. 32, no. 4, pp. 282-292, 2002.

[10] A. Nayak, "User centered approach for the design of knowledge discovery: systems used in technology management," *Portland International Conference on*

Management of Engineering and Technology, PICMET'99, vol. 1, p. 111, 1999.

[11] Rong Zhao and William I. Grosky, "Narrowing the Semantic Gap - Improved Text-Based Web Document Retrieval Using Visual Features," *IEEE Transactions on Multimedia*, vol. 4, pp. 189-200, 2002.

[12] Alfred Kobsa, Jürgen Koenemann, and Wolfgang Pohl, "Personalized Hypermedia Presentation Techniques for Improving Online Customer Relationships," *The Knowledge Engineering Review*, vol. 16, no. 2, pp. 111--155, 2001.

[13] Cormac Hampson and Owen Conlan, "Supporting Personalised Information Exploration through Subjective Expert-created Semantic Attributes," *3rd IEEE International Conference on Semantic Computing*, September 2009.

[14] Cormac Hampson, "Semantically Holistic and Personalized Views Across Heterogeneous Information Sources," *Workshop on Semantic Media Adaptation and Personalization, SMAP07, London, UK*, 17-18 December 2007.

[15] Cormac Hampson, "User-Centric Exploration of Heterogeneous Information Sources," Knowledge and Data Engineering Group (KDEG), School of Computer Science & Statistics, Trinity College Dublin (TCD), Dublin, Transfer Report 2008.

[16] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Transactions on Information Systems (TOIS)*, vol. 22, no. 1, pp. 5-53, 2004.

[17] Zan Huang, Hsinchun Chen, and Daniel Zeng, "Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering," *ACM Transactions on Information Systems (TOIS)*, vol. 22, no. 1, pp. 116-142, 2004.

[18] Gio Wiederhold, "Foreword: On the Barriers and Future of Knowledge Discovery," in *Advances in Knowledge Discovery and Data*, U. Fayyad et al., Eds.: AAAI/MIT Press, 1996, pp. 9-11.

[19] D. Mayer, A. Halevy, and M. Franklin, "Dataspaces: Coexistence with Heterogeneity," in *Tenth International Conference on Principles of Knowledge Representation and Reasoning*, Lake District of the United Kingdom, 2006, p. 3.

[20] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "The KDD process for extracting useful knowledge from volumes of data," *Communications of the ACM*, vol. 39, no. 11, pp. 27-34, 1996.

[21] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery: an overview," in *Advances in Knowledge Discovery and Data Mining*, U, Piatetsky-Shapiro, G, Smyth, P and Uthurusamy, R Fayyad, Ed.: AAAI Press, 1996, pp. 1-34.

[22] P. Cabena, P. Hadjinian, R. Stadler, and J Verhees, *Discovering Data Mining: From Concepts to Implementation*.: Prentice Hall, 1998.

[23] S. Anand and A. Buchner, *Decision Support Using Data Mining. Financial Time Management*. London, 1998.

[24] S. Anand, A. Patrick, J. Hughes, and D. Bell, "A data mining methodology for cross-sales," *Knowledge Based Systems Journal*, vol. 10, pp. 449-461, 1998.

[25] C. Shearer, "The CRISP-DM model: the new blueprint for data mining," *Journal of Data Warehousing*, vol. 15, no. 4, pp. 13-19, 2000.

[26] K. Cios, A. Teresinska, S. Konieczna, J. Potocka, and S. Sharma, "Diagnosing myocardial perfusion from PECT bull's-eye maps—a knowledge discovery approach," *IEEE Engineering in Medicine and Biology Magazine, Special issue on Medical Data Mining and Knowledge Discovery*, vol. 19, no. 4, pp. 17-25, 2000.

[27] Rayid Ghani and Andrew E. Fano, "Using Text Mining to Infer Semantic Attributes for Retail Data Mining," in *IEEE International Conference on Data Mining*, Maebashi City, Japan, 2002, pp. 195- 202.

[28] Suzanne Little, Joost Geurts, and Jane Hunter, "Dynamic Generation of Intelligent Multimedia Presentations through Semantic Inferencing," in *In 6th European Conference on Research and Advanced Technology for Digital Libraries*, Rome, 2002, pp. 158-189.

[29] E. Hyvönen, S. Saarela, and K. Viljanen, "Ontogator: combining view- and ontology-based search with semantic browsing," *Proceedings of XML*, 2003.

[30] E. Makela, E. Hyvonen, S. Saarela, and K. Viljanen, "OntoViews - A Tool for Creating Semantic Web Portals," *Lecture Notes in Computer Science*, no. 3298, pp. 797-811 , 2004.

[31] Cormac Hampson and Owen Conlan, "Supporting Personalised Information Exploration through Subjective Expert-created Semantic Attributes," Knowledge and Data Engineering Group (KDEG), Trinity College Dublin (TCD), Dublin, 2009.

[32] Alfred Kobsa, "User modeling and user-adapted interaction," in *Conference companion on Human factors in computing systems*, Boston, Massachusetts, United States , 1994, pp. 415 - 416.

[33] Peter Brusilovsky, "Methods and Techniques of Adaptive Hypermedia," *User Modeling and User Adapted Interaction*, vol. 6, no. 2-3, pp. 87-129, 1996.

[34] Dermot Browne, Peter Totterdell, and Mike Norman, Eds., *Adaptive User Interfaces*. London, UK, UK: Academic Press Ltd., 1990.

[35] Josef Fink, Alfred Kobsa, and Andreas Nill, "User-Oriented Adaptivity and Adaptability in the AVANTI Project," in *Designing for the Web: Empirical Studies*, Redmond, WA, 1996.

[36] David N. Chin, "Acquiring user models," *Artificial Intelligence Review*, vol. 7, no. 3-4, pp. 185-197, August 1993.

[37] Elaine Rich, "User modeling via stereotypes," *Cognitive Science*, vol. 3, no. 4, pp. 329-

354, January 1979.

[38] R. Kass and I. Stadnyk, "Using User Models to Improve Organizational Information," in *Proceedings of the 3rd International Workshop on User Modeling*, Dagstuhl, Germany, 1992.

[39] Judy Kay, "Scrutable Adaptation: Because We Can and Must," in *Proc. of 4th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH 2006)*, V. Wade, B. Smyth, and H. Ashman, Eds. Heidelberg: Springer Berlin, 2006, ch. 4018, pp. 11-19.

[40] Judy Kay, "Learner Control ," *User Modeling and User-Adapted Interaction*, vol. 11, no. 1-2, pp. 111-127, March 2001.

[41] (2009, August) IMDb. [Online]. http://www.imdb.com/chart/

[42] (2009, August) IMDb. [Online]. http://www.imdb.com/Sections/Recommendations/

[43] (2009, August) Amazon. [Online]. www.amazon.com

[44] (2009, August) GroupLens. [Online]. http://www.grouplens.org/

[45] (2009, August) What to Rent. [Online]. http://www.whattorent.com

[46] (2009, August) Criticker. [Online]. http://www.criticker.com

[47] (2009, August) Clerkdogs. [Online]. http://www.clerkdogs.com/

[48] (2009, August) Jinni. [Online]. http://www.jinni.com

[49] (2009, August) Nanocrowd. [Online]. http://www.nanocrowd.com/

[50] (2009, August) TasteKid. [Online]. http://www.tastekid.com/

[51] (2009, August) Flixter. [Online]. http://www.flixster.com/

[52] (2009, August) Netflix. [Online]. http://www.netflix.com/

[53] (2009, August) MovieProfiler. [Online]. http://www.movieprofiler.com/

[54] (2009, August) LinkedMDB. [Online]. http://www.linkedmdb.org/

[55] (2009, August) FindAnyFilm. [Online]. http://FindAnyFilm.com

[56] (2009, August) Spout. [Online]. http://www.spout.com/

[57] (2009, August) Cooliris. [Online]. http://www.cooliris.com

[58] (2009, August) Google Image Search. [Online]. http://images.google.com/

[59] (2009, August) YouTube. [Online]. http://www.youtube.com/

[60] (2009, August) Flickr. [Online]. http://www.flickr.com

[61] (2009, August) Newsmap. [Online]. http://newsmap.jp

[62] (2009, August) Google News. [Online]. http://news.google.com/

[63] L. Byron and M. Wattenberg, "Stacked graphs – geometry & aesthetics," *IEEE Trans. Vis. and Comp. Graphics*, vol. 14, no. 6, pp. 1245-1252, 2008.

[64] (2008, August) Last.fm. [Online]. http://www.last.fm

[65] (2009, August) Lastgraph3. [Online]. http://lastgraph3.aeracode.org

[66] (2009, August) Musicovery. [Online]. http://www.musicovery.com

[67] (2009, August) FIDGT. [Online]. http://www.fidgt.com

[68] Alan Cooper, Robert Reimann, and David Cronin, *About Face 3: The Essentials of Interaction Design*, 3rd ed. Indianapolis, IN: John Wiley & Sons, 2007.

[69] Sebastian Baumgärtner, Achim Ebert, Matthias Deller, and Stefan Agne, "2D meets 3D: a human-centered interface for visual data exploration," in *Conference on Human Factors in Computing Systems. CHI '07 extended abstracts on Human factors*

in computing systems, San Jose, CA, USA, 2007, pp. 2273 - 2278.

[70] (2009, September) Apache Tomcat. [Online]. http://tomcat.apache.org/

[71] (2009, September) eXist Open Source Native XML Database. [Online]. http://exist.sourceforge.net/

[72] (2009, September) XQuery. [Online]. www.w3.org/TR/xquery

[73] (2009, September) Adobe Flash Platform Technologies. [Online]. http://labs.adobe.com/technologies/flash/

[74] (2009, September) BlazeDS. [Online]. http://opensource.adobe.com/wiki/display/blazeds/BlazeDS

[75] (2009, September) Adobe Flex Framework Technologies. [Online]. http://labs.adobe.com/technologies/flex/

[76] (2009, September) Adobe AIR Technologies. [Online]. http://labs.adobe.com/technologies/air/

[77] (2009, September) Ecma International. [Online]. http://www.ecma-international.org/publications/standards/Ecma-357.htm

[78] (2009, September) Altova XMLSpy. [Online]. http://www.altova.com/xmlspy.html

[79] (2009, September) eclipse. [Online]. http://www.eclipse.org/

[80] (2009, September) SAX. [Online]. http://www.saxproject.org/

[81] (2009, September) Document Object Model. [Online]. http://www.w3.org/DOM/

[82] (2009, September) Adobe Photoshop. [Online]. http://www.adobe.com/products/photoshop

[83] (2009, September) Freebase - Film. [Online]. http://www.freebase.com/view/film

[84] (2009, September) New York Times Movies. [Online]. movies.nytimes.com

[85] (2009, August) Infoplease. [Online]. http://www.infoplease.com/ipa/A0777584.html

[86] (2009, September) Box Office Mojo. [Online]. http://www.boxofficemojo.com/

[87] (2009, September) Wikipedia. [Online]. http://www.wikipedia.org

[88] (2009, September) The Numbers. [Online]. http://www.the-numbers.com/

[89] (2009, September) Allmovie. [Online]. http://www.allmovie.com/

[90] (2009, September) VIM. [Online]. http://www.vim.org/

[91] (2009, September) Notepad++. [Online]. http://notepad-plus.sourceforge.net/uk/site.htm

[92] (2009, September) SlickEdit. [Online]. http://www.slickedit.com/

[93] (2009, September) Windows Grep. [Online]. http://www.wingrep.com

[94] (2009, September) MoviePosterDB. [Online]. http://www.movieposterdb.com/

[95] (2009, September) IMP Awards. [Online]. http://www.impawards.com/

[96] A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin, *Bayesian Data Analysis*, Second Edition ed., Chapman & Hall/CRC, Ed.: Physica Verlag, An Imprint of Springer-Verlag GmbH, 2004.

[97] S. Weibel, J. Kunze, C. Lagoze, and M. Wolf, "Dublin Core Metadata for Resource Discovery," *Internet Engineering Task Force RFC*, vol. 2413, September 1998.

[98] Cormac HHHampson and Owen Conlan, "Supporting Personalised Information Exploration through Subjective Expert-created Semantic Attributes," Knowledge and Data Engineering Group (KDEG), Trinity College Dublin (TCD), Dublin, 2009.

Appendix A: Abbreviations

AS3 ActionScript3

DOM Document Object Model

DSSP Dataspace Support System

FDEC Film Domain Exploration Client

GUI Graphical User Interface

HCI Human-computer interaction

IDE Integrated development environment

IMDb The Internet Movie Database

KDDM Knowledge Discovery and Data Mining

MPAA Motion Picture Association of America

RIA Rich Internet applications

RT Rotten Tomatoes

SA Semantic Attribute

SARA Semantic Attribute Reconciliation Architecture

SAX Simple API for XML

TCI Taste Compatibility Index

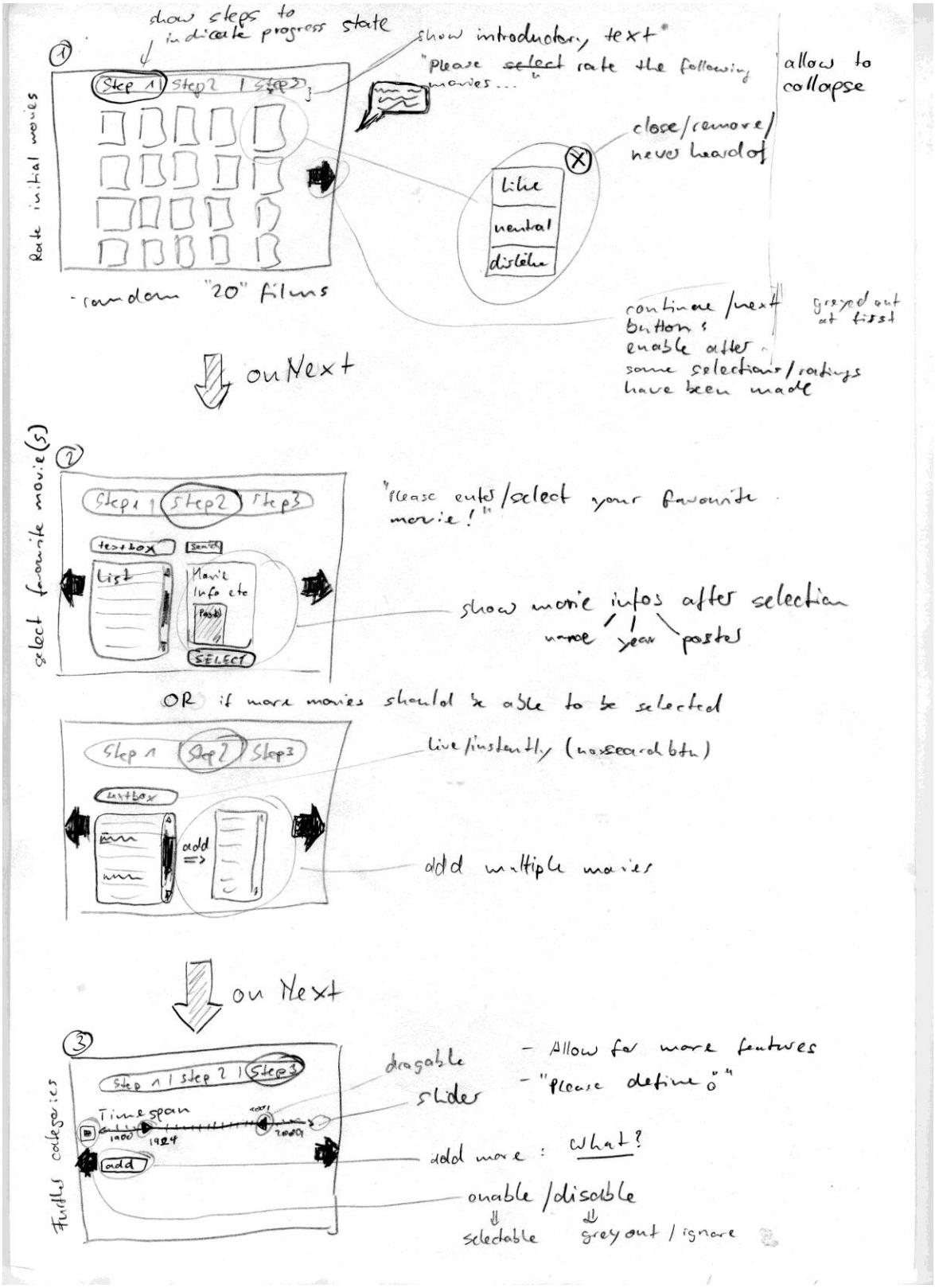
UI User Interface

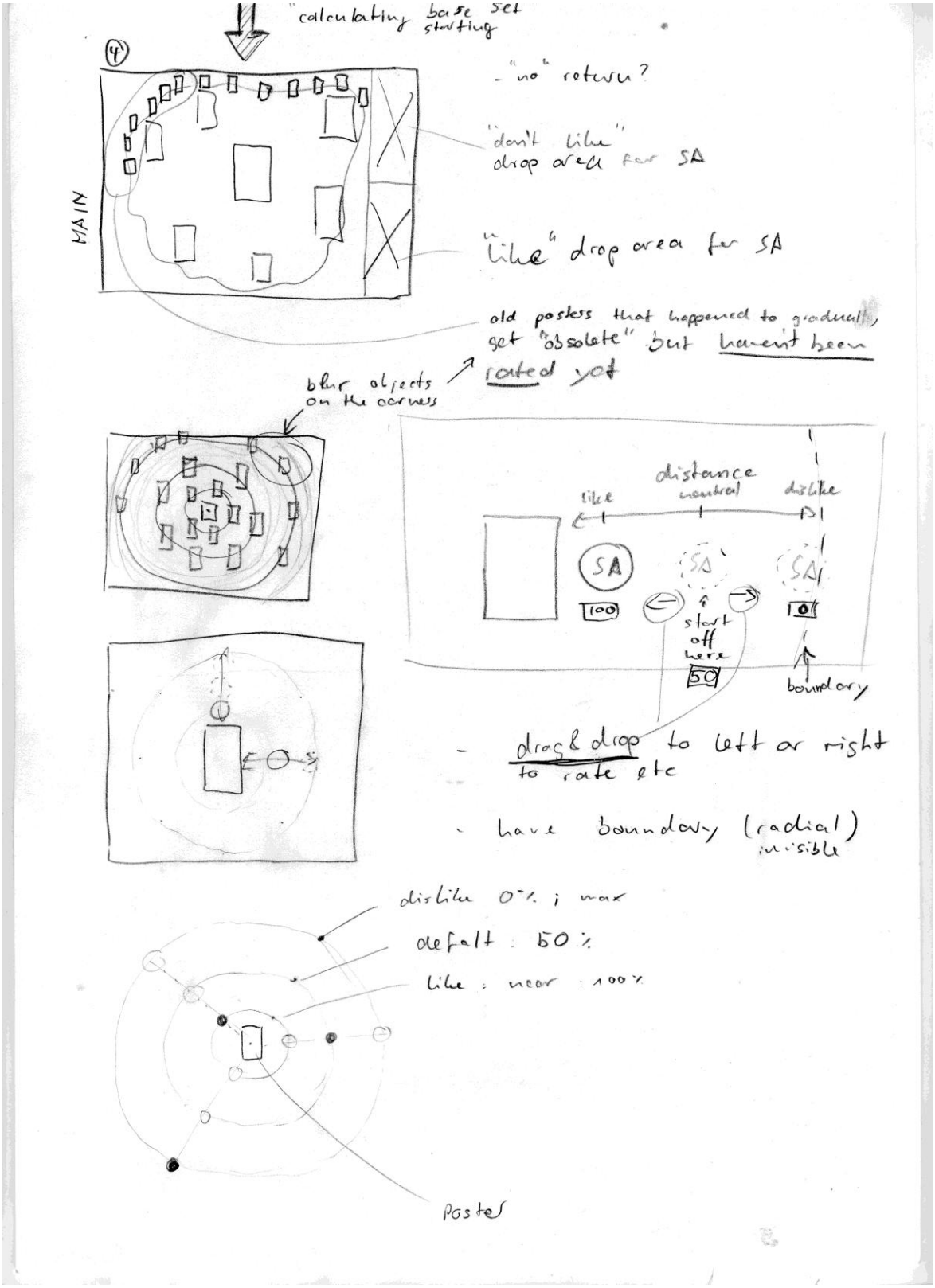
UM User Modelling

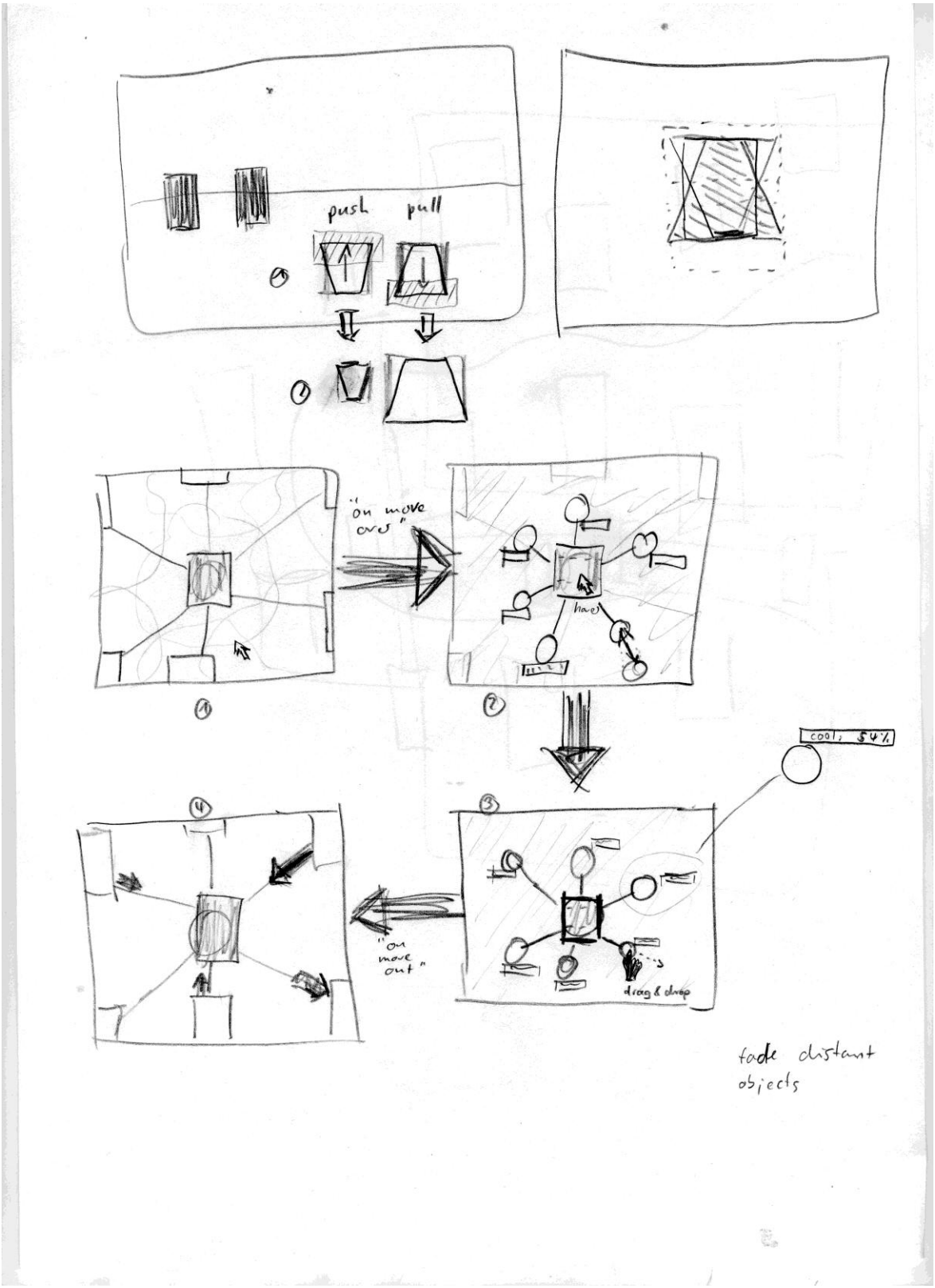
XML Extensible Markup Language

Appendix B: Film Domain Exploration Client screenshots

Appendix C: Sketches




"calculating base set starting
4
MAIN
- "no" return?
"don't like" drop area for SA
"like" drop area for SA
old posters that happened to gradually get "obsolete" but haven't been rated yet
blur objects on the corners
distance
like
neutral
dislike
SA
100
0
start off here
50
boundary
- drag & drop to left or right to rate etc
- have boundary (radial) invisible
dislike 0% ; max
default : 50 %
like : near : 100 %
Posters


push
pull
"on move over"
have
cool, $4%
"on move out"
drag & drop
fade distant objects

Appendix D: Dataset example


```
<movie>
	<title>Groundhog Day</title>
	<url>groundhogday</url>
	<rating_imdb>8.1</rating_imdb>
	<year>1993</year>
	<runtime>101</runtime>
	<budget>14600000</budget>
	<grossing>70906973</grossing>
	<grossing_worldwide>70906973</grossing_worldwide>
	<award_academy>false</award_academy>
	<award_bafta>false</award_bafta>
	<award_berlin>false</award_berlin>
	<award_cannes>false</award_cannes>
	<award_europeanfilmaward>false</award_europeanfilmaward>
	<award_goldenglobe>false</award_goldenglobe>
	<award_mtv>false</award_mtv>
	<award_slamdance>false</award_slamdance>
	<award_sundance>false</award_sundance>
	<award_toronto>false</award_toronto>
	<award_venice>false</award_venice>
	<director>Harold Ramis</director>
	<producer>Harold Ramis</producer>
	<producer>Trevor Albert</producer>
	<cinematographer>John Bailey</cinematographer>
	<actor>Andie MacDowell</actor>
	<actor>Brian Doyle-Murray</actor>
	<actor>Chris Elliott</actor>
	<actor>Bill Murray</actor>
	<actor>Stephen Tobolowsky</actor>
	<actor>Sandy Maschmeyer</actor>
	<actor>Marita Geraghty</actor>
	<actor>Angela Paton</actor>
	<actor>C.O. Erickson</actor>
	<actor>Chet Dubowski</actor>
	<actor>Rick Overton</actor>
	<actor>Lee R. Sellars</actor>
	<actor>David Pasquesi</actor>
	<actor>Rick Ducommun</actor>
	<actor>Robin Duke</actor>
	<actor>Carol Bivins</actor>
	<actor>Harold Ramis</actor>
	<actor>Peggy Roeder</actor>
	<actor>Willie Garson</actor>
	<actor>Ken Hudson Campbell</actor>
	<actor>John Watson Sr.</actor>
	<actor>Tom Milanovich</actor>
	<actor>Les Podewell</actor>
	<actor>Rod Sell</actor>
	<genre>Time travel</genre>
	<genre>Drama</genre>
	<genre>Comedy</genre>
	<genre>Fantasy</genre>
	<genre>Romance</genre>
</movie>
<movie>
	<title>Spider-man 3</title>
	<url>spiderman3</url>
	<rating_imdb>6.5</rating_imdb>
	<year>2007</year>
	<runtime>139</runtime>
	<budget>258000000</budget>
```

```
        <grossing>336530303</grossing>
        <grossing_worldwide>890871626</grossing_worldwide>
        <award_academy>false</award_academy>
        <award_bafta>false</award_bafta>
        <award_berlin>false</award_berlin>
        <award_cannes>false</award_cannes>
        <award_europeanfilmaward>false</award_europeanfilmaward>
        <award_goldenglobe>false</award_goldenglobe>
        <award_mtv>false</award_mtv>
        <award_slamdance>false</award_slamdance>
        <award_sundance>false</award_sundance>
        <award_toronto>false</award_toronto>
        <award_venice>false</award_venice>
        <director>Sam Raimi</director>
        <producer>Laura Ziskin</producer>
        <producer>Stan Lee</producer>
        <producer>Avi Arad</producer>
        <producer>Sony Pictures Entertainment</producer>
        <producer>Grant Curtis</producer>
        <cinematographer>Bill Pope</cinematographer>
        <actor>Stan Lee</actor>
        <actor>Tobey Maguire</actor>
        <actor>Kirsten Dunst</actor>
        <actor>James Franco</actor>
        <actor>Thomas Haden Church</actor>
        <actor>Topher Grace</actor>
        <actor>Bryce Dallas Howard</actor>
        <actor>J. K. Simmons</actor>
        <actor>Bruce Campbell</actor>
        <actor>Dylan Baker</actor>
        <actor>Cliff Robertson</actor>
        <actor>Elizabeth Banks</actor>
        <actor>Theresa Russell</actor>
        <actor>James Cromwell</actor>
        <actor>Bill Nunn</actor>
        <actor>Michael Papajohn</actor>
        <actor>Willem Dafoe</actor>
        <actor>Alfred Molina</actor>
        <actor>Rosemary Harris</actor>
        <genre>Sci-Fi</genre>
        <genre>Action</genre>
        <genre>Adventure</genre>
        <genre>Superhero</genre>
</movie>
<movie>
        <title>Quantum Of Solace</title>
        <url>jamesbond22</url>
        <rating_imdb>6.9</rating_imdb>
        <year>2008</year>
        <runtime>106</runtime>
        <budget>230000000</budget>
        <grossing>169368427</grossing>
        <grossing_worldwide>576368427</grossing_worldwide>
        <award_academy>false</award_academy>
        <award_bafta>false</award_bafta>
        <award_berlin>false</award_berlin>
        <award_cannes>false</award_cannes>
        <award_europeanfilmaward>false</award_europeanfilmaward>
        <award_goldenglobe>false</award_goldenglobe>
        <award_mtv>false</award_mtv>
        <award_slamdance>false</award_slamdance>
        <award_sundance>false</award_sundance>
        <award_toronto>false</award_toronto>
        <award_venice>false</award_venice>
        <director>Marc Forster</director>
        <producer>Michael G. Wilson</producer>
        <producer>Barbara Broccoli</producer>
        <cinematographer/>
```

```
        <actor>Judi Dench</actor>
        <actor>Jesper Christensen</actor>
        <actor>Giancarlo Giannini</actor>
        <actor>Jeffrey Wright</actor>
        <actor>Gemma Arterton</actor>
        <actor>Olga Kurylenko</actor>
        <actor>Daniel Craig</actor>
        <actor>Mathieu Amalric</actor>
        <actor>S. P. Balasubrahmanyam</actor>
        <genre>Action</genre>
        <genre>Adventure</genre>
</movie>
<movie>
        <title>Transformers: Revenge Of The Fallen</title>
        <url>transformers2</url>
        <rating_imdb>6.3</rating_imdb>
        <year>2009</year>
        <runtime>110</runtime>
        <budget>210000000</budget>
        <grossing>200077255</grossing>
        <grossing_worldwide>389596848</grossing_worldwide>
        <award_academy>false</award_academy>
        <award_bafta>false</award_bafta>
        <award_berlin>false</award_berlin>
        <award_cannes>false</award_cannes>
        <award_europeanfilmaward>false</award_europeanfilmaward>
        <award_goldenglobe>false</award_goldenglobe>
        <award_mtv>false</award_mtv>
        <award_slamdance>false</award_slamdance>
        <award_sundance>false</award_sundance>
        <award_toronto>false</award_toronto>
        <award_venice>false</award_venice>
        <director>Michael Bay</director>
        <producer>Steven Spielberg</producer>
        <producer>Don Murphy</producer>
        <producer>Lorenzo di Bonaventura</producer>
        <producer>Tom DeSanto</producer>
        <cinematographer/>
        <actor>Hugo Weaving</actor>
        <actor>Michael Papajohn</actor>
        <actor>Peter Cullen</actor>
        <actor>Josh Duhamel</actor>
        <actor>Shia LaBeouf</actor>
        <actor>Tyrese</actor>
        <actor>John Turturro</actor>
        <actor>Matthew Marsden</actor>
        <actor>Megan Fox</actor>
        <genre>Action</genre>
        <genre>Adventure</genre>
        <genre>Sci-Fi</genre>
</movie>
```

Appendix E: Modelling results

```
*****************************************************************************************
** New Session: Fri Sep 4 18:17:07 GMT+0100 2009
*****************************************************************************************
[1] 00:01:07: Initial Focus Movie: Miami Vice
[2] 00:01:07: Liked films (total, multiple rounds): 0
[3] 00:01:07: Neutral films (total, multiple rounds): 0
[4] 00:01:07: Disliked films (total, multiple rounds): 0
[5] 00:01:07: Total rated films: 0
[6] 00:01:07: PopularRating(liked, neutral, disliked): 0,0,0
[7] 00:01:07: PopularGrossingWorld(liked, neutral, disliked): 0,0,0
[8] 00:01:07: AwardWinner(liked, neutral, disliked): 0,0,0
[9] 00:01:07: PopularGrossingWorldBudgetRatio(liked, neutral, disliked): 0,0,0
[10] 00:01:07: Current Sem Atts (weighting): PopularRating: 5
[11] 00:01:07: Current Sem Atts (weighting): PopularGrossingWorld: 5
[12] 00:01:07: Current Sem Atts (weighting): AwardWinner: 5
[13] 00:01:07: Current Sem Atts (weighting): PopularGrossingWorldBudgetRatio: 5
[14] 00:01:07: Sem Att PopularRating (SARA query parameters): 5.4112343240000005,6.61124324
[15] 00:01:07: Sem Att PopularGrossingWorld (SARA query parameters): 131054845.01212001,196582267.411212
[16] 00:01:07: Sem Att AwardWinner (SARA query parameters):
false,false,false,false,false,false,false,false,false,false,false
[17] 00:01:07: Sem Att PopularGrossingWorldBudgetRatio (SARA query parameters):
0.8112424,1.2112412099999998

**** New focus movie ****************************************
[18] 00:01:28: Rating: 1 |  | Miami Vice
[19] 00:01:46: Current Focus Movie: Bad Boys II (PopularRating)
[20] 00:01:52: Liked films (total, multiple rounds): 1
[21] 00:01:52: Neutral films (total, multiple rounds): 0
[22] 00:01:52: Disliked films (total, multiple rounds): 0
[23] 00:01:52: Total rated films: 1
[24] 00:01:52: PopularRating(liked, neutral, disliked): 0,0,0
[25] 00:01:52: PopularGrossingWorld(liked, neutral, disliked): 0,0,0
[26] 00:01:52: AwardWinner(liked, neutral, disliked): 0,0,0
[27] 00:01:52: PopularGrossingWorldBudgetRatio(liked, neutral, disliked): 0,0,0
[28] 00:01:52: Current Sem Atts (weighting): PopularRating: 6
[29] 00:01:52: Current Sem Atts (weighting): PopularGrossingWorld: 6
[30] 00:01:52: Current Sem Atts (weighting): AwardWinner: 6
[31] 00:01:52: Current Sem Atts (weighting): PopularGrossingWorldBudgetRatio: 6
[32] 00:01:52: Sem Att PopularRating (SARA query parameters): 5.591234324,6.831243240000001
[33] 00:01:52: Sem Att PopularGrossingWorld (SARA query parameters): 218352696.21212,327529044.211212
[34] 00:01:52: Sem Att AwardWinner (SARA query parameters):
false,false,false,false,false,false,false,false,false,false,false
[35] 00:01:52: Sem Att PopularGrossingWorldBudgetRatio (SARA query parameters): 1.6112424,2.41124121

**** New focus movie ****************************************
[36] 00:02:12: Rating: 1 | PopularRating | Bad Boys II
[37] 00:02:23: Rating: 0 | PopularRating | Bad Boys II
[38] 00:02:43: Rating: 1 | PopularRating | Bad Boys II
[39] 00:03:03: Rating: 0 | PopularRating | Bad Boys II
[40] 00:03:10: Rating: 1 | PopularRating | Bad Boys II
[41] 00:03:31: Rating: 0 | PopularGrossingWorld | Gone In 60 Seconds
[42] 00:03:35: Rating: 1 | PopularGrossingWorld | Gone In 60 Seconds
[43] 00:03:48: Current Focus Movie: Gone In 60 Seconds (PopularGrossingWorld)
[44] 00:03:52: Liked films (total, multiple rounds): 3
[45] 00:03:52: Neutral films (total, multiple rounds): 0
[46] 00:03:52: Disliked films (total, multiple rounds): 0
[47] 00:03:52: Total rated films: 3
[48] 00:03:52: PopularRating(liked, neutral, disliked): 1,0,0
[49] 00:03:52: PopularGrossingWorld(liked, neutral, disliked): 1,0,0
[50] 00:03:52: AwardWinner(liked, neutral, disliked): 0,0,0
[51] 00:03:52: PopularGrossingWorldBudgetRatio(liked, neutral, disliked): 0,0,0
[52] 00:03:52: Current Sem Atts (weighting): PopularRating: 9
[53] 00:03:52: Current Sem Atts (weighting): PopularGrossingWorld: 9
[54] 00:03:52: Current Sem Atts (weighting): AwardWinner: 4
[55] 00:03:52: Current Sem Atts (weighting): PopularGrossingWorldBudgetRatio: 4
[56] 00:03:52: Sem Att PopularRating (SARA query parameters): 5.4112343240000005,6.61124324
[57] 00:03:52: Sem Att PopularGrossingWorld (SARA query parameters): 186114406.61212,279171609.81121194
[58] 00:03:52: Sem Att AwardWinner (SARA query parameters):
false,false,false,false,false,false,false,false,false,false,false
[59] 00:03:52: Sem Att PopularGrossingWorldBudgetRatio (SARA query parameters): 1.6112424,2.41124121

**** New focus movie ****************************************
[60] 00:04:05: Rating: 1 | PopularRating | Heist
[61] 00:04:25: Rating: 0 | PopularGrossingWorldBudgetRatio | Lions For Lambs
[62] 00:05:09: Current Focus Movie: Heist (PopularRating)
[63] 00:05:13: Liked films (total, multiple rounds): 4
[64] 00:05:13: Neutral films (total, multiple rounds): 1
[65] 00:05:13: Disliked films (total, multiple rounds): 0
[66] 00:05:13: Total rated films: 4
[67] 00:05:13: PopularRating(liked, neutral, disliked): 2,0,0
[68] 00:05:13: PopularGrossingWorld(liked, neutral, disliked): 1,0,0
[69] 00:05:13: AwardWinner(liked, neutral, disliked): 0,0,0
[70] 00:05:13: PopularGrossingWorldBudgetRatio(liked, neutral, disliked): 0,1,0
[71] 00:05:13: Current Sem Atts (weighting): PopularRating: 13
```

```
[72] 00:05:13: Current Sem Atts (weighting): PopularGrossingWorld: 8
[73] 00:05:13: Current Sem Atts (weighting): AwardWinner: 4
[74] 00:05:13: Current Sem Atts (weighting): PopularGrossingWorldBudgetRatio: 4
[75] 00:05:13: Sem Att PopularRating (SARA query parameters): 5.771234324000001,7.051243240000001
[76] 00:05:13: Sem Att PopularGrossingWorld (SARA query parameters): 22786534.612120003,34179801.811212
[77] 00:05:13: Sem Att AwardWinner (SARA query parameters):
false,false,false,false,false,false,false,false,false,false,false
[78] 00:05:13: Sem Att PopularGrossingWorldBudgetRatio (SARA query parameters): 0.0112424,1.01124121

**** New focus movie ****************************************
[79] 00:05:24: Rating: 0 | PopularRating | The Score
[80] 00:05:37: Rating: 1 | PopularRating | The Score
[81] 00:05:55: Rating: 1 | PopularRating | Hitman
[82] 00:06:03: Current Focus Movie: Hitman (PopularRating)
[83] 00:06:09: Liked films (total, multiple rounds): 6
[84] 00:06:09: Neutral films (total, multiple rounds): 1
[85] 00:06:09: Disliked films (total, multiple rounds): 0
[86] 00:06:09: Total rated films: 6
[87] 00:06:09: PopularRating(liked, neutral, disliked): 4,0,0
[88] 00:06:09: PopularGrossingWorld(liked, neutral, disliked): 1,0,0
[89] 00:06:09: AwardWinner(liked, neutral, disliked): 0,0,0
[90] 00:06:09: PopularGrossingWorldBudgetRatio(liked, neutral, disliked): 0,1,0
[91] 00:06:09: Current Sem Atts (weighting): PopularRating: 16
[92] 00:06:09: Current Sem Atts (weighting): PopularGrossingWorld: 6
[93] 00:06:09: Current Sem Atts (weighting): AwardWinner: 3
[94] 00:06:09: Current Sem Atts (weighting): PopularGrossingWorldBudgetRatio: 3
[95] 00:06:09: Sem Att PopularRating (SARA query parameters): 5.681234324,6.94124324
[96] 00:06:09: Sem Att PopularGrossingWorld (SARA query parameters): 79972633.81212,119958950.61121199
[97] 00:06:09: Sem Att AwardWinner (SARA query parameters):
false,false,false,false,false,false,false,false,false,false,false
[98] 00:06:09: Sem Att PopularGrossingWorldBudgetRatio (SARA query parameters): 4.0112424,6.01124121

**** New focus movie ****************************************
[99] 00:06:40: Rating: 1 | PopularRating | The Fast And The Furious
[100] 00:07:00: Current Focus Movie: Home Alone 2: Lost In New York (PopularRating)
[101] 00:07:04: Liked films (total, multiple rounds): 7
[102] 00:07:04: Neutral films (total, multiple rounds): 1
[103] 00:07:04: Disliked films (total, multiple rounds): 0
[104] 00:07:04: Total rated films: 7
[105] 00:07:04: PopularRating(liked, neutral, disliked): 5,0,0
[106] 00:07:04: PopularGrossingWorld(liked, neutral, disliked): 1,0,0
[107] 00:07:04: AwardWinner(liked, neutral, disliked): 0,0,0
[108] 00:07:04: PopularGrossingWorldBudgetRatio(liked, neutral, disliked): 0,1,0
[109] 00:07:04: Current Sem Atts (weighting): PopularRating: 17
[110] 00:07:04: Current Sem Atts (weighting): PopularGrossingWorld: 5
[111] 00:07:04: Current Sem Atts (weighting): AwardWinner: 2
[112] 00:07:04: Current Sem Atts (weighting): PopularGrossingWorldBudgetRatio: 2
[113] 00:07:04: Sem Att PopularRating (SARA query parameters): 5.141234324,6.28124324
[114] 00:07:04: Sem Att PopularGrossingWorld (SARA query parameters): 287195880.21212,430793820.211212
[115] 00:07:04: Sem Att AwardWinner (SARA query parameters):
false,false,false,false,false,false,false,false,false,false,false
[116] 00:07:04: Sem Att PopularGrossingWorldBudgetRatio (SARA query parameters):
13.611242400000002,20.41124121

**** New focus movie ****************************************
[117] 00:07:29: Rating: 1 | PopularRating | Collateral Damage
[118] 00:07:39: Current Focus Movie: Captain Corelli's Mandolin (PopularRating)
[119] 00:07:48: Liked films (total, multiple rounds): 8
[120] 00:07:48: Neutral films (total, multiple rounds): 1
[121] 00:07:48: Disliked films (total, multiple rounds): 0
[122] 00:07:48: Total rated films: 8
[123] 00:07:48: PopularRating(liked, neutral, disliked): 6,0,0
[124] 00:07:48: PopularGrossingWorld(liked, neutral, disliked): 1,0,0
[125] 00:07:48: AwardWinner(liked, neutral, disliked): 0,0,0
[126] 00:07:48: PopularGrossingWorldBudgetRatio(liked, neutral, disliked): 0,1,0
[127] 00:07:48: Current Sem Atts (weighting): PopularRating: 18
[128] 00:07:48: Current Sem Atts (weighting): PopularGrossingWorld: 5
[129] 00:07:48: Current Sem Atts (weighting): AwardWinner: 2
[130] 00:07:48: Current Sem Atts (weighting): PopularGrossingWorldBudgetRatio: 2
[131] 00:07:48: Sem Att PopularRating (SARA query parameters): 5.141234324,6.28124324
[132] 00:07:48: Sem Att PopularGrossingWorld (SARA query parameters): 50022796.21212,75034194.211212
[133] 00:07:48: Sem Att AwardWinner (SARA query parameters):
false,false,false,false,false,false,false,false,false,false,false
[134] 00:07:48: Sem Att PopularGrossingWorldBudgetRatio (SARA query parameters):
0.8112424,1.2112412099999998

**** New focus movie ****************************************
[135] 00:08:21: Rating: 1 | PopularRating | Captain Corelli's Mandolin
[136] 00:09:01: Current Focus Movie: Knockaround Guys (PopularRating)
[137] 00:09:05: Liked films (total, multiple rounds): 9
[138] 00:09:05: Neutral films (total, multiple rounds): 1
[139] 00:09:05: Disliked films (total, multiple rounds): 0
[140] 00:09:05: Total rated films: 9
[141] 00:09:05: PopularRating(liked, neutral, disliked): 7,0,0
[142] 00:09:05: PopularGrossingWorld(liked, neutral, disliked): 1,0,0
[143] 00:09:05: AwardWinner(liked, neutral, disliked): 0,0,0
[144] 00:09:05: PopularGrossingWorldBudgetRatio(liked, neutral, disliked): 0,1,0
[145] 00:09:05: Current Sem Atts (weighting): PopularRating: 19
```

```
[146] 00:09:05: Current Sem Atts (weighting): PopularGrossingWorld: 5
[147] 00:09:05: Current Sem Atts (weighting): AwardWinner: 2
[148] 00:09:05: Current Sem Atts (weighting): PopularGrossingWorldBudgetRatio: 2
[149] 00:09:05: Sem Att PopularRating (SARA query parameters): 5.321234324000001,6.501243240000001
[150] 00:09:05: Sem Att PopularGrossingWorld (SARA query parameters): 9935760.21212,14903640.211212
[151] 00:09:05: Sem Att AwardWinner (SARA query parameters):
false,false,false,false,false,false,false,false,false,false,false
[152] 00:09:05: Sem Att PopularGrossingWorldBudgetRatio (SARA query parameters): 0.0112424,1.01124121

**** New focus movie ****************************************
[153] 00:09:14: Current Focus Movie: Birthday Girl (PopularRating)
[154] 00:09:18: Liked films (total, multiple rounds): 9
[155] 00:09:18: Neutral films (total, multiple rounds): 1
[156] 00:09:18: Disliked films (total, multiple rounds): 0
[157] 00:09:18: Total rated films: 9
[158] 00:09:18: PopularRating(liked, neutral, disliked): 7,0,0
[159] 00:09:18: PopularGrossingWorld(liked, neutral, disliked): 1,0,0
[160] 00:09:18: AwardWinner(liked, neutral, disliked): 0,0,0
[161] 00:09:18: PopularGrossingWorldBudgetRatio(liked, neutral, disliked): 0,1,0
[162] 00:09:18: Current Sem Atts (weighting): PopularRating: 19
[163] 00:09:18: Current Sem Atts (weighting): PopularGrossingWorld: 5
[164] 00:09:18: Current Sem Atts (weighting): AwardWinner: 2
[165] 00:09:18: Current Sem Atts (weighting): PopularGrossingWorldBudgetRatio: 2
[166] 00:09:18: Sem Att PopularRating (SARA query parameters): 5.4112343240000005,6.61124324
[167] 00:09:18: Sem Att PopularGrossingWorld (SARA query parameters): 6504581.812120001,9756872.611212
[168] 00:09:18: Sem Att AwardWinner (SARA query parameters):
false,false,false,false,false,false,false,false,false,false,false
[169] 00:09:18: Sem Att PopularGrossingWorldBudgetRatio (SARA query parameters): 0.0112424,1.01124121

**** New focus movie ****************************************
[170] 00:09:39: Current Focus Movie: The Night Listener (PopularRating)
[171] 00:09:48: Liked films (total, multiple rounds): 9
[172] 00:09:48: Neutral films (total, multiple rounds): 1
[173] 00:09:48: Disliked films (total, multiple rounds): 0
[174] 00:09:48: Total rated films: 9
[175] 00:09:48: PopularRating(liked, neutral, disliked): 7,0,0
[176] 00:09:48: PopularGrossingWorld(liked, neutral, disliked): 1,0,0
[177] 00:09:48: AwardWinner(liked, neutral, disliked): 0,0,0
[178] 00:09:48: PopularGrossingWorldBudgetRatio(liked, neutral, disliked): 0,1,0
[179] 00:09:48: Current Sem Atts (weighting): PopularRating: 19
[180] 00:09:48: Current Sem Atts (weighting): PopularGrossingWorld: 5
[181] 00:09:48: Current Sem Atts (weighting): AwardWinner: 2
[182] 00:09:48: Current Sem Atts (weighting): PopularGrossingWorldBudgetRatio: 2
[183] 00:09:48: Sem Att PopularRating (SARA query parameters): 5.321234324000001,6.501243240000001
[184] 00:09:48: Sem Att PopularGrossingWorld (SARA query parameters): 8438204.21212,12657306.211212
[185] 00:09:48: Sem Att AwardWinner (SARA query parameters):
false,false,false,false,false,false,false,false,false,false,false
[186] 00:09:48: Sem Att PopularGrossingWorldBudgetRatio (SARA query parameters): 1.6112424,2.41124121

**** New focus movie ****************************************
[187] 00:10:03: Rating: 1 | PopularRating | The Night Listener
[188] 00:10:56: Current Focus Movie: Antitrust (PopularGrossingWorld)
[189] 00:11:00: Liked films (total, multiple rounds): 10
[190] 00:11:00: Neutral films (total, multiple rounds): 1
[191] 00:11:00: Disliked films (total, multiple rounds): 0
[192] 00:11:00: Total rated films: 10
[193] 00:11:00: PopularRating(liked, neutral, disliked): 8,0,0
[194] 00:11:00: PopularGrossingWorld(liked, neutral, disliked): 1,0,0
[195] 00:11:00: AwardWinner(liked, neutral, disliked): 0,0,0
[196] 00:11:00: PopularGrossingWorldBudgetRatio(liked, neutral, disliked): 0,1,0
[197] 00:11:00: Current Sem Atts (weighting): PopularRating: 20
[198] 00:11:00: Current Sem Atts (weighting): PopularGrossingWorld: 4
[199] 00:11:00: Current Sem Atts (weighting): AwardWinner: 2
[200] 00:11:00: Current Sem Atts (weighting): PopularGrossingWorldBudgetRatio: 2
[201] 00:11:00: Sem Att PopularRating (SARA query parameters): 5.4112343240000005,6.61124324
[202] 00:11:00: Sem Att PopularGrossingWorld (SARA query parameters):
8772167.412120001,13158251.011211999
[203] 00:11:00: Sem Att AwardWinner (SARA query parameters):
false,false,false,false,false,false,false,false,false,false,false
[204] 00:11:00: Sem Att PopularGrossingWorldBudgetRatio (SARA query parameters): 0.0112424,1.01124121

**** New focus movie ****************************************
[205] 00:11:08: Rating: 1 | PopularGrossingWorld | Antitrust
[206] 00:11:29: Current Focus Movie: High Crimes (PopularRating)
[207] 00:11:33: Liked films (total, multiple rounds): 11
[208] 00:11:33: Neutral films (total, multiple rounds): 1
[209] 00:11:33: Disliked films (total, multiple rounds): 0
[210] 00:11:33: Total rated films: 11
[211] 00:11:33: PopularRating(liked, neutral, disliked): 8,0,0
[212] 00:11:33: PopularGrossingWorld(liked, neutral, disliked): 2,0,0
[213] 00:11:33: AwardWinner(liked, neutral, disliked): 0,0,0
[214] 00:11:33: PopularGrossingWorldBudgetRatio(liked, neutral, disliked): 0,1,0
[215] 00:11:33: Current Sem Atts (weighting): PopularRating: 18
[216] 00:11:33: Current Sem Atts (weighting): PopularGrossingWorld: 6
[217] 00:11:33: Current Sem Atts (weighting): AwardWinner: 2
[218] 00:11:33: Current Sem Atts (weighting): PopularGrossingWorldBudgetRatio: 2
[219] 00:11:33: Sem Att PopularRating (SARA query parameters): 5.501234324,6.72124324
[220] 00:11:33: Sem Att PopularGrossingWorld (SARA query parameters): 51024880.21212,76537320.211212
```

```
[221] 00:11:33: Sem Att AwardWinner (SARA query parameters):
false,false,false,false,false,false,false,false,false,false,false
[222] 00:11:33: Sem Att PopularGrossingWorldBudgetRatio (SARA query parameters):
0.8112424,1.2112412099999998

**** New focus movie ****************************************
[223] 00:11:49: Rating: -1 | PopularRating | All The Pretty Horses
[224] 00:12:09: Rating: -1 | PopularRating | The Other Side Of Heaven
[225] 00:12:14: Rating: -1 | PopularRating | Hoodwinked
[226] 00:12:18: Rating: 1 | PopularRating | Behind Enemy Lines
[227] 00:12:23: Rating: 1 | PopularRating | Double Jeopardy
[228] 00:12:28: Rating: 1 | PopularRating | High Crimes
[229] 00:12:41: Rating: -1 | PopularRating | Keeping The Faith
[230] 00:12:46: Rating: -1 | PopularRating | Armageddon
[231] 00:12:55: Current Focus Movie: Double Jeopardy (PopularRating)
[232] 00:12:59: Liked films (total, multiple rounds): 14
[233] 00:12:59: Neutral films (total, multiple rounds): 1
[234] 00:12:59: Disliked films (total, multiple rounds): 5
[235] 00:12:59: Total rated films: 19
[236] 00:12:59: PopularRating(liked, neutral, disliked): 11,0,5
[237] 00:12:59: PopularGrossingWorld(liked, neutral, disliked): 2,0,0
[238] 00:12:59: AwardWinner(liked, neutral, disliked): 0,0,0
[239] 00:12:59: PopularGrossingWorldBudgetRatio(liked, neutral, disliked): 0,1,0
[240] 00:12:59: Current Sem Atts (weighting): PopularRating: 17
[241] 00:12:59: Current Sem Atts (weighting): PopularGrossingWorld: 7
[242] 00:12:59: Current Sem Atts (weighting): AwardWinner: 2
[243] 00:12:59: Current Sem Atts (weighting): PopularGrossingWorldBudgetRatio: 2
[244] 00:12:59: Sem Att PopularRating (SARA query parameters): 5.4112343240000005,6.61124324
[245] 00:12:59: Sem Att PopularGrossingWorld (SARA query parameters): 142268185.01212,213402277.411212
[246] 00:12:59: Sem Att AwardWinner (SARA query parameters):
false,false,false,false,false,false,false,false,false,false,false
[247] 00:12:59: Sem Att PopularGrossingWorldBudgetRatio (SARA query parameters):
3.2112424,4.8112412099999995

**** New focus movie ****************************************
[248] 00:13:28: Rating: -1 | PopularRating | Tomorrow Never Dies
[249] 00:13:31: Rating: 0 | PopularRating | Tomorrow Never Dies
[250] 00:13:33: Rating: 1 | PopularRating | Tomorrow Never Dies
[251] 00:13:39: Rating: 1 | PopularRating | The Sum Of All Fears
[252] 00:13:44: Rating: -1 | PopularRating | The Princess Diaries
[253] 00:13:48: Rating: -1 | PopularRating | Analyze That
[254] 00:13:51: Rating: -1 | PopularRating | Laws Of Attraction
[255] 00:13:55: Rating: -1 | PopularGrossingWorld | Stuart Little 2
[256] 00:13:59: Current Focus Movie: Tomorrow Never Dies (PopularRating)
[257] 00:14:03: Liked films (total, multiple rounds): 16
[258] 00:14:03: Neutral films (total, multiple rounds): 1
[259] 00:14:03: Disliked films (total, multiple rounds): 9
[260] 00:14:03: Total rated films: 25
[261] 00:14:03: PopularRating(liked, neutral, disliked): 13,0,8
[262] 00:14:03: PopularGrossingWorld(liked, neutral, disliked): 2,0,1
[263] 00:14:03: AwardWinner(liked, neutral, disliked): 0,0,0
[264] 00:14:03: PopularGrossingWorldBudgetRatio(liked, neutral, disliked): 0,1,0
[265] 00:14:03: Current Sem Atts (weighting): PopularRating: 17
[266] 00:14:03: Current Sem Atts (weighting): PopularGrossingWorld: 5
[267] 00:14:03: Current Sem Atts (weighting): AwardWinner: 2
[268] 00:14:03: Current Sem Atts (weighting): PopularGrossingWorldBudgetRatio: 2
[269] 00:14:03: Sem Att PopularRating (SARA query parameters): 5.771234324000001,7.051243240000001
[270] 00:14:03: Sem Att PopularGrossingWorld (SARA query parameters): 271603421.01212,407405131.41121197
[271] 00:14:03: Sem Att AwardWinner (SARA query parameters):
false,false,false,false,false,false,false,false,false,false,false
[272] 00:14:03: Sem Att PopularGrossingWorldBudgetRatio (SARA query parameters):
2.4112424000000003,3.6112412099999998

**** New focus movie ****************************************
[273] 00:14:21: Rating: -1 | PopularRating | Nurse Betty
[274] 00:14:26: Rating: -1 | PopularRating | Thirteen
[275] 00:14:31: Rating: -1 | PopularRating | Pooh's Heffalump Movie
[276] 00:14:36: Rating: 1 | PopularRating | Derailed
[277] 00:14:52: Rating: -1 | PopularRating | The Upside Of Anger
[278] 00:15:11: Rating: -1 | PopularRating | Hamlet 2
[279] 00:15:16: Rating: -1 | PopularRating | The Statement
[280] 00:15:47: Rating: 1 | PopularGrossingWorld | True Lies
[281] 00:15:50: Current Focus Movie: True Lies (PopularGrossingWorld)
[282] 00:15:56: Liked films (total, multiple rounds): 18
[283] 00:15:56: Neutral films (total, multiple rounds): 1
[284] 00:15:56: Disliked films (total, multiple rounds): 15
[285] 00:15:56: Total rated films: 33
[286] 00:15:56: PopularRating(liked, neutral, disliked): 14,0,14
[287] 00:15:56: PopularGrossingWorld(liked, neutral, disliked): 3,0,1
[288] 00:15:56: AwardWinner(liked, neutral, disliked): 0,0,0
[289] 00:15:56: PopularGrossingWorldBudgetRatio(liked, neutral, disliked): 0,1,0
[290] 00:15:56: Current Sem Atts (weighting): PopularRating: 4
[291] 00:15:56: Current Sem Atts (weighting): PopularGrossingWorld: 15
[292] 00:15:56: Current Sem Atts (weighting): AwardWinner: 4
[293] 00:15:56: Current Sem Atts (weighting): PopularGrossingWorldBudgetRatio: 4
[294] 00:15:56: Sem Att PopularRating (SARA query parameters): 6.4912343240000006,7.931243240000001
[295] 00:15:56: Sem Att PopularGrossingWorld (SARA query parameters): 292240000.21212,438360000.211212
```

```
[296] 00:15:56: Sem Att AwardWinner (SARA query parameters):
false,false,false,false,false,false,false,false,false,false,false
[297] 00:15:56: Sem Att PopularGrossingWorldBudgetRatio (SARA query parameters):
2.4112424000000003,3.6112412099999998

**** New focus movie ****************************************
[298] 00:16:11: Current Focus Movie: Grease (PopularGrossingWorld)
[299] 00:16:17: Liked films (total, multiple rounds): 18
[300] 00:16:17: Neutral films (total, multiple rounds): 1
[301] 00:16:17: Disliked films (total, multiple rounds): 15
[302] 00:16:17: Total rated films: 33
[303] 00:16:17: PopularRating(liked, neutral, disliked): 14,0,14
[304] 00:16:17: PopularGrossingWorld(liked, neutral, disliked): 3,0,1
[305] 00:16:17: AwardWinner(liked, neutral, disliked): 0,0,0
[306] 00:16:17: PopularGrossingWorldBudgetRatio(liked, neutral, disliked): 0,1,0
[307] 00:16:17: Current Sem Atts (weighting): PopularRating: 4
[308] 00:16:17: Current Sem Atts (weighting): PopularGrossingWorld: 15
[309] 00:16:17: Current Sem Atts (weighting): AwardWinner: 4
[310] 00:16:17: Current Sem Atts (weighting): PopularGrossingWorldBudgetRatio: 4
[311] 00:16:17: Sem Att PopularRating (SARA query parameters): 6.311234324,7.711243240000001
[312] 00:16:17: Sem Att PopularGrossingWorld (SARA query parameters): 309766808.21212,464650212.211212
[313] 00:16:17: Sem Att AwardWinner (SARA query parameters):
false,false,false,false,false,false,false,false,false,false,false
[314] 00:16:17: Sem Att PopularGrossingWorldBudgetRatio (SARA query parameters):
51.2112424,76.81124120999999

**** New focus movie ****************************************
[315] 00:17:39: Rating: 1 | PopularGrossingWorld | Superman Returns
[316] 00:17:43: Rating: 1 | PopularGrossingWorld | Cast Away
[317] 00:17:46: Current Focus Movie: Cast Away (PopularGrossingWorld)
[318] 00:17:50: Liked films (total, multiple rounds): 20
[319] 00:17:50: Neutral films (total, multiple rounds): 1
[320] 00:17:50: Disliked films (total, multiple rounds): 15
[321] 00:17:50: Total rated films: 35
[322] 00:17:50: PopularRating(liked, neutral, disliked): 14,0,14
[323] 00:17:50: PopularGrossingWorld(liked, neutral, disliked): 5,0,1
[324] 00:17:50: AwardWinner(liked, neutral, disliked): 0,0,0
[325] 00:17:50: PopularGrossingWorldBudgetRatio(liked, neutral, disliked): 0,1,0
[326] 00:17:50: Current Sem Atts (weighting): PopularRating: 3
[327] 00:17:50: Current Sem Atts (weighting): PopularGrossingWorld: 19
[328] 00:17:50: Current Sem Atts (weighting): AwardWinner: 3
[329] 00:17:50: Current Sem Atts (weighting): PopularGrossingWorldBudgetRatio: 3
[330] 00:17:50: Sem Att PopularRating (SARA query parameters): 6.671234324,8.151243240000001
[331] 00:17:50: Sem Att PopularGrossingWorld (SARA query parameters): 341784413.01212,512676619.41121197
[332] 00:17:50: Sem Att AwardWinner (SARA query parameters):
false,false,false,false,false,false,false,false,false,false,false
[333] 00:17:50: Sem Att PopularGrossingWorldBudgetRatio (SARA query parameters): 4.0112424,6.01124121

**** New focus movie ****************************************
[334] 00:18:01: Rating: 1 | PopularGrossingWorld | Ocean's Twelve
[335] 00:18:05: Rating: 1 | PopularGrossingWorld | Top Gun
[336] 00:18:08: Current Focus Movie: Top Gun (PopularGrossingWorld)
[337] 00:18:14: Liked films (total, multiple rounds): 22
[338] 00:18:14: Neutral films (total, multiple rounds): 1
[339] 00:18:14: Disliked films (total, multiple rounds): 15
[340] 00:18:14: Total rated films: 37
[341] 00:18:14: PopularRating(liked, neutral, disliked): 14,0,14
[342] 00:18:14: PopularGrossingWorld(liked, neutral, disliked): 7,0,1
[343] 00:18:14: AwardWinner(liked, neutral, disliked): 0,0,0
[344] 00:18:14: PopularGrossingWorldBudgetRatio(liked, neutral, disliked): 0,1,0
[345] 00:18:14: Current Sem Atts (weighting): PopularRating: 2
[346] 00:18:14: Current Sem Atts (weighting): PopularGrossingWorld: 20
[347] 00:18:14: Current Sem Atts (weighting): AwardWinner: 2
[348] 00:18:14: Current Sem Atts (weighting): PopularGrossingWorldBudgetRatio: 2
[349] 00:18:14: Sem Att PopularRating (SARA query parameters): 5.861234324000001,7.16124324
[350] 00:18:14: Sem Att PopularGrossingWorld (SARA query parameters): 283029361.01212,424544041.41121197
[351] 00:18:14: Sem Att AwardWinner (SARA query parameters):
false,false,false,false,false,false,false,false,false,false,false
[352] 00:18:14: Sem Att PopularGrossingWorldBudgetRatio (SARA query parameters):
18.411242400000003,27.61124121
```

Appendix F: Evaluation questionnaire

Personal

1.	Name	.....
2.	Age	.....
3.	Gender	Male ☐ Female ☐
4.	How comfortable are you with computers?	Not at all ☐ A little ☐ Somewhat ☐ Moderately ☐ Quite a lot ☐ Very much ☐

Films in general

5.	I am interested in films.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
6.	I know a lot about films.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
7.	I watch more films now compared to 10 years ago.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
8.	I watch a lot of films.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
9.	I regularly get information about films from various sources.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
10.	Where do you usually hear about new films?	On the radio ☐ On TV ☐ In the cinema ☐ In magazines/ newspapers ☐ Online ☐			

		Don't know ☐ Other .....
11.	How often do you use online services or websites to get information about films?	Very often ☐ Often ☐ Rarely ☐ Very rarely ☐ Not at all ☐
12.	Please select any of the following film related websites you have **heard of**!	IMDb ☐ Freebase ☐ Rotten Tomatoes ☐ MovieLens ☐ What To Rent ☐ Criticker ☐ Clerkdogs ☐ Jinni ☐ Other .....
13.	Please select any of the following film related websites you have **used at least once**!	IMDb ☐ Freebase ☐ Rotten Tomatoes ☐ MovieLens ☐ What To Rent ☐ Criticker ☐ Clerkdogs ☐ Jinni ☐ Other .....
14.	Please select any of the following film related websites you **use regularly!**	IMDb ☐ Freebase ☐ Rotten Tomatoes ☐ MovieLens ☐ What To Rent ☐ Criticker ☐ Clerkdogs ☐ Jinni ☐ Other .....
15.	When browsing for films, what types of information are you interested in most?	General information ☐ New films ☐

		Upcoming films ☐ Actors ☐ Popular films ☐ Films from specific genre ☐ Don't know ☐ Other			
16.	I like "Blockbuster films".	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
17.	I like "Expensive films".	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
18.	I like "Award winning films".	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
19.	I like "Successful films".	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
20.	I like...	.....			

Software client / UI

21.	It is easy to select an initial film with the tool I just used.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
22.	It is clear that I can change the initial film before I continue.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
23.	It is easy to rate films.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
24.	Setting a new focus film is simple to do.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
25.	It is intuitive to zoom.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
26.	Panning (navigating the stage vertically and horizontally) is easy.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
27.	It is easy to get lost.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐

28.	Showing the film poster image is useful.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
29.	Showing the film title on the poster helps indentifying the films.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
30.	It is simple to discover films I previously liked/disliked.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
31.	The user interface makes it easy to explore films.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
32.	The user interface is fun to use.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐

Further questions / general statements

33.	It was difficult to come up with an initial film in the first step when using the program.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
34.	I found films I like.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
35.	I found films I expected the software will show to me.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
36.	I think the system worked well.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
37.	The system was predictable / it is clear what is going on *"under the hood".*	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
38.	Do you remember how often you refocused on a new film?	Yes ☐ Number: ..... No ☐			
39.	Exploring films this way is a waste of time.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
40.	I am happy with the films that have been shown to me.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
41.	I wasn't aware that I know that many films.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐

42.	I was shown films I haven't been thinking about for a long time.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
43.	If I had time right now I would like to watch one of the films I found.	I strongly agree ☐	I agree ☐	I disagree ☐	I strongly disagree ☐
44.	What do you think the system is good for? What can it be used for?	.....			
45.	What would you change? (Navigation, Rating, Visualisation...)	.....			
46.	What do you think are the most significant remaining usability issues?	.....			
47.	What do you like most/least about the film exploring tool?	.....			
48.	How could the system be improved overall? Other features?	.....			
49.	Any other comments, ideas, questions?	.....			