Proc. of the IEEE 4th Int. Database Engineering and Application Symposium (IDEAS 2000),
18.-20. Sept. 2000, Yokohama, Japan, IEEE Computer Society, 2000, pp. 16-24

Information Filtering and Personalisation in Databases using Gaussian Curves

Günther Specht
Technische Universität München
Institut für Informatik
Orleansstr. 34, D-81667 München, Germany
specht@in.tum.de

Thomas Kahabka
Exolution GmbH
Barerstr. 60, D-80799 München, Germany
thomas.kahabka@exolution.de

Abstract

We present an information filtering and adaptive personalisation algorithm for arbitrary information systems based on databases. This algorithm is called GRAS (Gaussian Rating Adaptation Scheme), and it combines content-based and collaborative filtering. The goal is to filter retrieved documents of a query according to the personal interest of a user and to sort them according to the personal relevance. The algorithm tries to make the benefits of collaborative filtering available to application domains where collaborative filtering could not yet be applied due to lack of the critical mass of users or improper content structure. The algorithm collects background information about the user and the content by implicit and explicit feedback techniques. This information is then used to consecutively adapt user- and object profiles according their maturity. The described algorithm is applicable for the personalisation of any kind of application domain, even on multimedia data. GRAS is implemented in the multimedia database MultiMAP[1] as a generic personalisation provider module.

1 Introduction

We are living in an age of information overflow. The volume of information and the number of information sources are continuously increasing. In today's world, it is almost impossible to find and acquire relevant pieces of information without being overflowed with irrelevant material. With personalisation methods we can improve the situation.

An important technique used in personalised systems is information filtering which can be divided into three basic classes: content-based, collaborative and economic filtering. The wide research activities in information filtering have resulted in several content-based and collaborative filtering systems. Most of them have been built for one special application and rely on a certain type of content. This prevents them from being used as a generic personalisation scheme in a hypermedia database.

We needed to find a generic personalisation scheme for the multimedia database MultiMAP, which was developed in our research group at the Munich University of Technology. This lead to the Gaussian Rating Adaptation Scheme (GRAS) which combines content-based and collaborative filtering. Its efficient implementation using relational database technology allows online personalisation of hyperlinked multimedia objects. It is appropriate for the personalisation of multimedia and hyperlinked content, since it makes no assumptions about the structure of the objects, i.e. objects do not have to be text. It is not dependent on a critical mass of users or objects to work effectively, as many collaborative filtering approaches do, which widens the area of use for GRAS. The rest of this paper is organized as follows: In section 2 we give a short introduction to personalisation techniques. Section 3 presents the GRAS algorithm and in section 4 an efficient implementation using relational database technology is discussed. Section 5 gives a brief overview of MultiMAP, the host hypermedia database system used to demonstrate GRAS. Finally in section 6 we discuss related work.

2 Personalisation in Information Systems

The aim of personalisation is to select data whose content are most relevant to the user from a greater volume of information and to present them in a suitable way for the user.

The main logical parts (Fig 1) in a personalisation system are the user (and object) profile, user feedback and information filtering.

[1]MultiMAP is funded by the "Deutsches Forschungsnetz Verein" (DFN) and the "Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie" (BMBF) under contract TK 598-VA/V1.

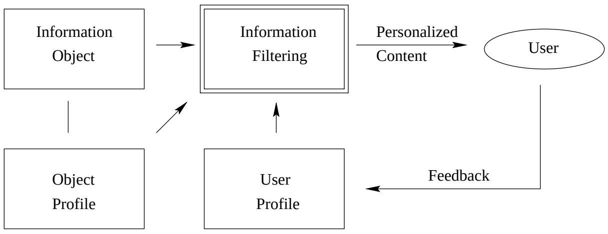

Figure 1. The structure of a personalisation system

2.1 User and Object Profiles

Information about each user's preferences (the user profile) must be available in order to select relevant data. The user profile can be constructed explicitly by the user or abstracted by the system on the basis of the user's behaviour. A third, more difficult approach is that the profile is defined and maintained by an administrator.
Some personalisation methods also make use of object profiles, which describe the contents or their characteristics. They are constructed explicitly by the creator, or implicitly by the system or the users. GRAS uses both user and object profiles. Known features can be defined explicitly while unknown profiles will be implicitly adapted by the GRAS algorithm.

2.2 Feedback

To determine if the data provided really satisfy the user, personalisation schemes often ask the user to give feedback, which is used to modify the user's profile.

Two different feedback methods exist: explicit and implicit. Explicit feedback requires the user to explicitly evaluate the provided objects. This scheme is easy to implement but increases the cognitive load on the user. Systems using implicit feedback observe the user's behaviour, like which content is read or how much time is spent on different content. Implicit feedback is more convenient for the user but harder to implement. The GRAS algorithm relies on user feedback and can make use of both feedback methods.

2.3 Information filtering

Information filtering is the main process in personalisation. It is traditionally divided into three types of filtering [11]:

Content-based filtering, also called cognitive filtering, where objects are selected by correlation between the content of the objects and the user's preferences. This is the traditional approach to filtering. Objects must be of some machine parsable form, or attributes must have been assigned to them manually. Hypermedia data such as sound, images, video, cannot be analysed automatically. It is not practical to assign attributes manually when working with a large object set. This makes content-based filtering less feasible for large hypermedia applications.

Collaborative filtering, also called social filtering, where objects are filtered for a user upon the likes of other people with similar tastes. The content of the objects need not to be in a computer parsable form which makes collaborative filtering applicable for hypermedia data. Originally, Malone [11], defined that social filtering takes place regarding social connections between people. Today, in many current systems a variation is used: user profiles are compared with each other and weighted for their degree of similarity. Groups of similar profiles are formed and users belonging to one group will be served the same set of objects. Collaborative filtering systems need a critical mass of participants and objects to work efficiently which is their major draw-back.

Economic filtering, where the information is filtered based upon cost factors. Such factors can be the relation between cost and benefit of use, or the available network bandwidth and size of the objects. Economic filtering is scarcely used in existing systems.

GRAS is a combination of content-based and collaborative filtering.

3 The GRAS Algorithm

GRAS was developed for the personalisation of hypermedia data which is not restricted to text. The algorithm should not be application dependant since its host hypermedia-database MultiMAP is used in a wide area of different applications. Additional requirements were efficiency for online-personalisation and ease of use for the user and author. A combination of content-based and collaborative filtering combined with a new model for user and object profiles were developed for achieving these requirements.

3.1 Profile Modelling

GRAS uses both user and object profiles. Each profile consists of several topic-profiles. Each topic-profile de-

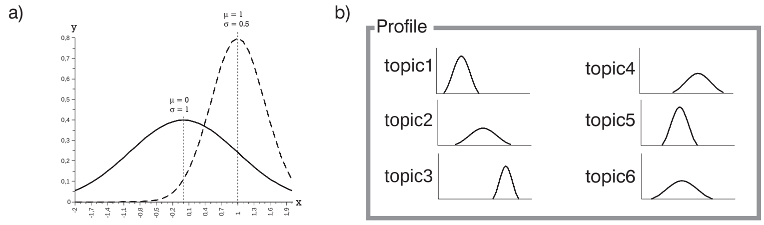

Figure 2. a) Examples of two different Gauss curves b) example of a profile consisting of several topic-profiles

scribes the interest of a user or the content of an object in a single topic. Examples of topics are news-categories like politics, sports or business.

3.1.1 Conventional ratings schemes compared to GRAS

Simple rating schemes [4, 16] use only one rating value to describe the user's interest in a topic. In these schemes, a high rating value means "the user is highly interested in the topic", while a low value means "the user is not interested in the topic". A one dimensional numeric scale may not be well suited to describe the reactions humans have to documents [13]. Works such as [20] and [12] show that persons have difficulty in expressing their interest explicitly on a single numeric scale.
These schemes often do not distinguish between "the user dislikes a topic" and "the user is indifferent to a topic". They also force the user to express his or her interest as one point on a rating-scale, giving no means to express the broadness of the interest.
In GRAS we are using two values to describe the user's interest: One value to describe the focus or median (μ) of the user's interest in a scale from "is highly interested in the topic" to "totally dislikes the topic" and a second value to describe the broadness (σ) of the user's interest.
To be able to work efficiently with this scheme, we are using Gaussian curves to represent the topic-profiles. A Gaussian curve is also characterized by two parameters μ and σ and is, as we later show, highly appropriate to describe the interest of a person or the content characteristics of an object. Figure 2 a) shows two Gauss curves with different μ and σ values.

The dotted curve can be interpreted as the interest profile of a highly interested person. A high, positive μ-value combined with a low σ-value standing for high significance represents interest only in objects whose content cover a specific topic intensively. The other curve represents the interest-profile of a user who is neither very interested in the topic nor dislikes it. Thus it has a μ-value of zero and a σ-value of one, which stands for not significant.
Figure 2 b) shows how several topic-profiles, each represented by a Gauss curve, form one object or user profile.

3.1.2 Considerations GRAS is based on

Consideration (1): The interest of a user in a topic can be described using a Gaussian curve. The x-axis represents the intensity in which an object covers a topic. The y-axis describes how much a user will be interested in an object. One can now plot a Gaussian curve, characterized by the parameters μ and σ, which describes the interest of a user in one topic. μ can be seen as a value describing how much a user is interested in the topic, while σ measures how broad the user's interest is.
Example: Figures 3 (a) and (b) show two different user profiles. The profile (a) shows a user who is indifferent to the specific topic thus having a profile with a high σ-value which makes it "broad". Profile (b) belongs to a user who is highly interested only in objects which are dealing intensively with the topic. This leads to a low σ-value, which makes the curve narrow, and a high μ-value for a high focus-point.

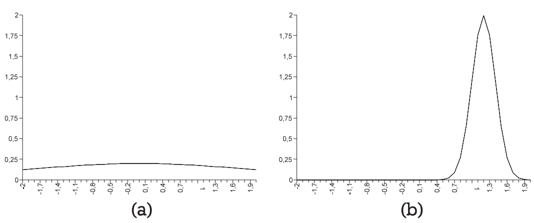

Figure 3. Examples of two different topic-profiles

Consideration (2): The profile of an object, describing how intensive an object covers a specific topic can also

be described using a Gaussian curve. The x-axis is used to describe the intensity of a topic. The y-axis now describes, how strongly an object covers different topic intensities. The parameter μ determines how intensively an object covers a topic, while σ tells how broadly the object's content covers a topic.
Example: Figures 3 (a) and (b) can also be used to explain two different topic-profiles of objects. (a) describes an object that covers the topic very broadly. (b) shows the topic-profile of an object whose content is specialized in the regarded topic thus resulting in a high μ- and a low σ-value.

These (3): A measure for how much a user is interested in an object regarding one topic is the overlap of the area of both Gaussian curves belonging to the user's profile and the object's profile.
We like to show that this gives a good measure by discussing how this method behaves in all extreme situations and argue that this covers all in-between possibilities as well. Figure 4 illustrates all example cases.

Situations A1 and C3: The object meets exactly the user's interest in a topic. The two Gaussian curves have a large area in common which creates a high interest value. Situation A1 is used only for negatable topics (i.e. "I don't like Jazz music") where it is positive if an object does not cover a topic and a user dislikes the topic. The decision models (see next section) have to take care about this special case.

Situations A3 and C1: The object is contrary to the user's interest in a topic. The Gaussian curves have a low to zero area in common which creates a low interest value.

Situation B1, B2 and B3: The object's content is indifferent to a topic.
According to consideration (2) this will cause the σ-value of the object's topic profile to be high. This will cause the area in common with any user's topic-profile in this specific topic to be independent of its μ-value. This corresponds with the desired behaviour, since the decision whether a user is interested in this object or not is not dependent on this specific topic.

Situations A2, B2 and C2: The user is indifferent to a topic.
Thus, according to consideration (1) the σ-value of the user's topic profile is high. This will cause the area in common with any object's topic-profile to be small, independent of the μ-value of the object. This means, that the decision whether a user would be interested in this object or not is independent of this topic.
Situation B2 shows that objects which are indifferent to the topic cause a high overlap and are therefore preferred. This means that a user who is indifferent to a topic prefers to see objects which are not specialized in this topic.

We can see, that the GRAS algorithm behaves as desired in all shown cases. In section 4 we present an efficient method of how to implement the overlap-calculation using relational database technology.

3.1.3 Decision models used in GRAS

We can now derive a measure of how much a user will like an object regarding each topic separately. That gives us a n-tupel of topic-overlap-values for an object. We need a model that describes how to compare or order these tuples to reach a decision about which object the user will like. One way is to convert the n-tupel to one single value and then compare them. Several algorithms, originating in marketing, describe how a consumer will make a decision based on multiple factors. They are known as "decision-models": The **additive model** [3], where the sum of all topic-ratings in the n-tupel is used to compare objects. The object with the highest sum will be chosen. The **conjunctive model** [21], where the lowest topic-rating of the n-tupel must be above a threshold to be acceptable. The **disjunctive model** [9], where the highest topic-rating of a n-tupel must be above a threshold to be acceptable.
We favour a variant of the additive model, since it takes all topic-overlaps into consideration to reach a decision. Our variant compares the average of the tupel, allowing the comparison of differently sized tupels. Aditionally, the decision model must take care, that a positive recommendation is not solely based on negative topic overlaps (situation A1 in figure 4).

3.2 Profile adaptation schemes

3.2.1 Profile generation in conventional filtering schemes

Most content-based filtering schemes are based on the assumption, that the user can tell exactly what he likes or dislikes. The user has to define his profile for the system. Collaborative filtering systems follow the opposite approach: they take assumptions from data gathered implicitly or explicitly about the user. This can be biographic data or sample ratings.
Both approaches can cause problems: it can be hard for the user to tell exactly his preferences in the system rating scheme. The average casual user is not willing to undergo a long profile-definition phase before using the system. Once

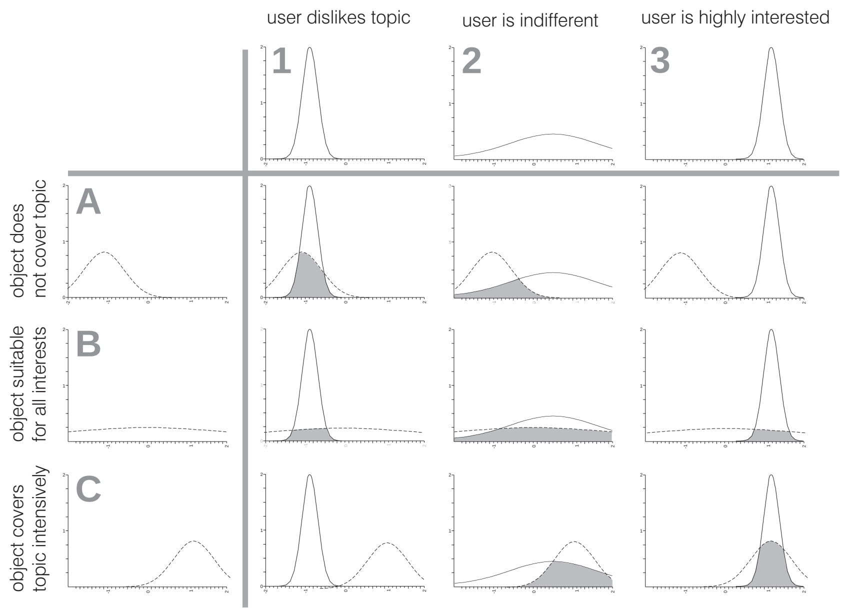

Figure 4. Overlap of object's profiles with user's profiles in GRAS

a profile is defined, it is even more unlikely that it will be changed by the user when only slight interest changes occur.

On the other side, pure collaborative filtering approaches which do not allow the explicit definition of preferences can also create problems. It is often a frustrating task to undergo a training-procedure to adapt the system to one's preferences. These systems also do not allow fast switches in preferences, for example when the user's interest changes in one area of interest. The biggest disadvantage of collaborative filtering schemes is that they require a critical mass of users which makes them unsuitable for a lot of applications.

3.2.2 Profile generation and update in GRAS

GRAS gives the user the possibility to explicitly define preferences he is sure about. At the same time it autonomously tries to find the true values for topic-profiles the user is uncertain about. This means, a user can explicitly tell the system, that he or she is highly interested in a specific topic or that another topic is extremely disliked. Symmetrically this can be applied to the object profiles. Authors can define topic-profiles for an object where they are sure about the rating. GRAS then tries to complete the profiles in the remaining topics.

For this task, each topic-profile contains, additional to the parameters describing the Gaussian curve, two parameters: an age value and a mode flag. The mode flag allows the topic-profile to be locked, which means it will not be part of the dynamic adaptation process. This is useful for topic rankings which may never change. For example, a movie-description will always have a high ranking in the topic "cinema-related". The age value, which stands for the maturity of the topic-profile and not for its time age, is used for the dynamic adaptation process. It counts, how often a not locked profile has been adjusted in an adaptation-step to become more mature.

3.2.3 The adaptation process in GRAS

Each time a user views an object the system collects an implicit or explicit feedback value, which describes how much the user liked the object. Upon receiving the rating, the system starts the adaptation process for the user's and the object's profile.

The dynamic adaptation process uses the ratio of the age values of the user's and the object's profile to determine whether the object's profile influences the user's profile or vice versa. The influenced profile's age value is incremented after each dynamic adaptation. The older profile will influence the younger profile. The strength of the influence is proportional to the age value ratio. This ensures that "mature" profiles with a higher "quality" will influence younger profiles and will increase their quality.

Beside the age-ratio, the strength of adaption also depends on the σ-value of the older profile. A highly significant profile (which has a low σ value) will influence the opposite profile more than a profile with a low significance. Another

variable that influences the strength of the adaptation is the user's feedback value, which has to be normalized to the interval [-1,1]. A feedback-value of 0 means "no influence", +1 means that the μ-values will move towards another and -1 that they will move apart from each other.
Only profiles of objects that are not marked as "locked" will be changed by the adaptation process. This means, if the "younger" profile in the adaptation process has the lock-flag set, the adaptation will not be conducted. It is no problem if the older profile is locked, since it will not be changed in the adaptation process.

3.3 Effects

The following effects can be achieved with the described adaptation scheme:

- Unknown values of profiles, which were initialized to a standard value (i.e. $\mu = 0, \sigma = 1$) will evolve to their true values. Objects which cover one topic intensively will be rated positive by users having high μ-values and low σ values in the respective topic-profile. Users with low μ-values in the specific topic-profile will rate this object negatively. This will lead to a highly significant positive topic-profile. Objects which do not cover a topic are handled symmetrically. In the same way immature user-profiles are influenced by mature object profiles making them more significant.
- The topic-profile of an object which is indifferent to the specific topic will evolve to a profile with a high σ-value. Users with all possible μ- and σ-values will judge this object independently of this topic thus creating a topic-profile with a high σ-value.
- Shifts in the interest of a user can be detected by the system when the user starts to rate objects covering a new topic of interest positively. Similarly newly arising dislikes of a user can be discovered.
- Users with similar interests will rate objects similarely. Thus the GRAS adaptation process will create similar user-profiles for these users. Similarely, users with a mature profile will influence object profiles thus allowing the system to propose objects which they find interesting to users with similar interest. That means, in GRAS' adaptation step an implicit grouping of users according to their interests takes place.

4 An Efficient Implementation in relational Databases

Some complex personalisation schemes suffer from poor performance. We want to provide real-time personalisation which can be used with live-data sources or material that has a high update frequency as well as with static data.

We use the following simplifications for the overlap-calculation:

- for calculating the overlap we reduce the number of possible values of μ and σ to m and n discrete values thus giving

$$\mu' = \lfloor \frac{\mu - \mu_{min}}{\mu_{max} - \mu_{min}} m \rfloor \qquad \sigma' = \lfloor \frac{\sigma - \sigma_{min}}{\sigma_{max} - \sigma_{min}} n \rfloor$$

- By choosing m and n to be 15, we can pack μ' and σ' into one tinyint (8 bit) b: $b = 16\mu' + \sigma' - 127$
- μ_{min} and μ_{max} define the interval of valid μ-values. Experiments showed that the interval $[-1.5, +1.5]$ is a good choice.
- σ_{min} and σ_{max} define the interval of valid σ values. Experiments showed that the interval $]0, 3]$ is a good choice.

When calculating the overlap, we use the one-byte representation of the object's topic-profile and the user's interest topic-profile. We can now create a relation "GaussLookup" containing three values as shown in table 1.

We fill the table GaussLookup with all possible permutations of b_1 and b_2. v then contains the overlap-value of the two Gaussian curves represented by the packed values b_1 and b_2. This table needs only 256kb (256*256*4 bytes), which is less than the size of many images that are stored in the multimedia database.
The user's and the object's topic-profile table are defined similarly. One relation "ObjProfile" for object's topic-profiles and one relation "UserProfile" for user's object-profiles are defined as shown in table 2.

To calculate how much a user's topic-profile overlaps with an object's topic profile we only have to use two joins over the three tables "UserProfile", "GaussLookup" and "ObjProfile":

```
select v, u.TopicID from GaussLookup g,
                    UserProfile u,
                    ObjProfile o
where u.UserID = USERID and
      o.ObjID = OBJECTID and
      u.TopicID = o.TopicID and
      g.b1 = u.bMuSigma and
      g.b2 = o.bMuSigma
```

This gives the list of the overlap-values between all the topic-profiles of the user identified by USERID and the

Name	Type	Function
b1	tinyint	the first packed (μ, σ)-value
b2	tinyint	the second packed (μ, σ)-value
v	integer	the overlap of the Gaussian curves represented by b_1 and b_2

Table 1. Scheme of the table GaussLookup

Name	Type	Function
UserID / ObjID	Integer	identifies the user or the object
TopicID	integer	identifies the topic of the rating
mu	double	the μ-value of the topic-profile
sigma	double	the σ-value of the topic-profile
bMuSigma	tinyint	the packed representation of μ and σ
age	integer	the age (maturity) of the topic-profile
locked	boolean	locked or not locked topic-profile

Table 2. Scheme of the tables UserProfile and ObjProfile

topic-profiles of the corresponding objects, identified by OBJID.

We can now use the methods described in section 3.1.3 to determine how much the user will overall like the object. Some of the described methods can also be implemented in the same query. Here an example for the additive model:

```
select sum(v) from GaussLookup g,
              UserProfile u, ObjProfile o
where u.UserID = USERID and
      o.ObjID = OBJECTID and
      u.TopicID = o.TopicID and
      g.b1 = u.bMuSigma and
      g.b2 = o.bMuSigma
```

Using other functions instead of `sum(v)` can vary the decision model: To work according to the disjunctive-model we can use `max(v)` and for the conjunctive model we use `min(v)`. Another interesting method is to use `avg(v)`, a variation of the additive method.

We use the result of the query to present the objects to the user in a ranked, pre-ordered way which allows fast access to potentially interesting objects.

The computational complexity of the whole operation is even below the complexity of two simple joins because the first join creates a constant number of elements, which is the number of existing topics. This reduces the complexity to $O(n \log n \log m)$.

The lookup-table requires approximately as much space as one image in the multimedia-database uses: size $= 256^2(1 + 1 + 2)$bytes $= 256$kb where 256^2 is the number of permutations of b_1 and b_2. Each tupel uses one byte each for `b1` and `b2`, which are tinyint, and two bytes for `v`, which is shortint. 256kb can easily be handled by a relational database.

The advantage of referring the task of overlap-calculation to the database more than compensates the disadvantage that `bMuSigma` stores redundant information. All insert/update operations must be designed to keep `bMuSigma` consistent with `mu` and `sigma`. One could argue, that `mu` and `sigma` are not necessary and could be eliminated since they are already stored in `bMuSigma`. During the adaptation process very small changes in `mu` and `sigma` are expected, which would then not be possible due to the packed representation of μ and σ. Since `bMuSigma` is declared as tinyint, the space trade-off is affordable.

5 The Hypermedia Database System MultiMAP

5.1 The System

GRAS is implemented as a generic personalisation module in MultiMAP. MultiMAP is an interactive, extensible hypermedia database system, in which texts, images, arbitrary objects on the images, audio and videos can be stored and connected by links. MultiMAP runs on Unix workstations using a client/server architecture, and the relational database system TransBase^{TM} as back-end for internal data management. MultiMAP is also accessible via a WWW interface.

One of the main focus of MultiMAP is the support of fast and simple (mouse supported) creation of applications. Due to its database functionality even deletion of nodes does

not touch the referential integrity of links. Thus always a consistent application is presented. Further advantages of using a database instead of the still very common file systems in multimedia systems are: integrated processing of big amounts of multimedia data (All multimedia data is completely stored inside the database. There are no pointers to the file system for BLOBs), optimized storage due to efficient access paths and index structures, multiple complex search possibilities, referential integrity of links, transaction protected multi-user mode and full recovery capability.
MultiMAP is conceptually based on an extension of the Dexter Hypertext Reference-Model [8, 19], which is an acknowledged standard for hypermedia systems today. This is important for the power of the link concept which goes far beyond usual WWW-links:

1. Support of uni- and bi-directional links and arbitrary n:m links. This includes the heavily used 1:n links in our applications.

2. Extension of the hypertext concept on arbitrary graphical objects: Link source and target anchors can be arbitrarily outlined objects on images (e.g. the course of a river or a plot of land on a map). These do not need to be approximated by rectangles.

3. In addition to links, it is possible to execute full text search (even truncated and nested) on all text, image and object names of the database. The full text search is integrated in the object recherche and behaves like an additional dynamic link.

In addition to complex link navigation full-text search is supported as an entry to the hypermedia net.

5.2 Applications

MultiMAP is already used in a series of applications, partially with large amounts of data and high user activity. We present only a few of them:

The first field of application was the multimedia processing of maps for urban information systems, mapping out biotopes, or administrative domains for environmental planning.
MultiMED: This application presents X-ray images in medicine, including detail images and verbal or written medical reports. A prototype has been developed in collaboration with the St. Bernward hospital in Hildesheim, Germany [18].
MultiBHT: A third field of application lies in linguistics, in multimedia processing of results of language analysis, in order to develop text-critical editions. An application for Old-Hebrew exists in the Institute for Assyriology and Hethitology of the University of Munich.
MultiLIB is a multimedia guide through the university library of the University of Munich. The purpose is to enable the students to find books, their location and access rights, opening times of the library, and to offer support in catalogue queries.

In each application a huge amount of documents is already inserted and a personalized retrieval using GRAS becomes more and more necessary for further efficient use.

6 Related Work

The interest in personalisation methods has continuously increased in the recent years. This is due to the easy accessibility of networked information via popular systems like the WWW.
Personalisation based on content-based filtering has a strong tradition in Mail and News filtering. SIFT is an example of a simple content-based text-filtering system for Internet News [22]. In SIFT the profile is based on words to prefer and avoid. InfoScope is another system designed to filter Internet News. It uses automatic profile learning based on reading behaviour of the user [20].
With the introduction of the World Wide Web, new areas applicable for personalisation emerged. One area is personalisation of electronic newspapers. Most methods use simple content-based filtering based on numerical and boolean rating schemes or word-matching techniques. An example is the personalisation scheme for an electronic newspaper developed in the OtaOnline Project at the Helsinki University of Technology [16]. Other examples are the Fishwrap at MIT [1], The Times[3] and The Sunday Times[4].
The Tapestry text filtering system, developed to filter Mail and News articles, was the first to include collaborative filtering [4]. The Internet News Filter GroupLens[5] is another system based on collaborative filtering [15]. GroupLens annotations are explicit judgements on a five-valued integer scale. Similar techniques were used in the Ringo personalised music recommendations system, developed at the MIT Media Lab [17]. The music and movie recommendation service Firefly[6] has evolved from the work done at the MIT Media Lab and Ringo. The catalogue service Yahoo also deploys a personalised service called my-Yahoo[7] which is said to be based on technology used in Firefly.
Other systems applying collaborative information filtering are the video recommendation service implemented at Bellcore [5], the Beehive developed at Xerox Palo Alto [7] and the News filtering system Phoaks [6]. Most of the used

[3]http://www.the-times.co.uk, [4]http://www.sunday-times.co.uk, [5]http://www.cs.umn.edu/Research/GroupLens, [6]http://www.firefly.com, [7]http://www.my- yahoo.com

content-based filtering methods are based on word matching and on the frequency with which terms occur. Two common approaches for content selection are the vector space method and the probabilistic method [13]. Other approaches try to include semantic information. Examples are: Latent Semantic Indexing (LSI) [2], natural language processing and the use of neural networks. An interesting example of the use of neural networks can be found in [10]. For collaborative filtering algorithms [17] is a good starting point.

7 Conclusion

Combining content-based and collaborative filtering can create a new personalisation scheme which avoids the shortcomings of conventional filtering systems.
With GRAS we present such an algorithm. It is based on Gaussian curves, which can describe user's interest and object's contents in a more distinguished way. The overlap of an object's profile-curve and a user's interest curve tells how much a user is interested in an object. Relational database-technology is used to efficiently implement the GRAS personalisation-scheme. A demonstrational implementation in the hypermedia database MultiMAP leads to first experiences with the GRASS algorithm.

We thank Michael G. Bauer for useful comments and suggestions.

References

[1] P. E. Chesnais, et al.: *The Fishwrap Personalized News System*, MIT Media Laboratory. http://fishwrap-docs.www.media.mit.edu/docs/dev/CNGlue/cnglue.html

[2] Nicholas DeClaris, et al.: *Information Filtering and Retrieval: Overview, Issues and Directions*, 26th Conference of the IEEE Engineering in Medicine and Biology Society, Baltimore, MD, November 3-6, 1994. http://www.ee.umd.edu/medlab/filter/papers/balt.ps

[3] W.L. Etter: *Attitude Theory and Decision Theory: Where Is the Common Ground?*, Journal of Marketing Research (JMR) Vol. 12, 1975, pp. 481-483.

[4] D. Goldberg, et al.: *Using Collaborative Filtering to Waeve an Information Tapestry*, Communications of the ACM, 1992, Vol. 25 No. 12, pp. 61-70.

[5] W. Hill, et al.: *Recommending and Evaluating Choices in a Virtual Community of Use*, Bellcore Labs. http://community.bellcore.com/navigation/video.html

[6] W. Hill and L. Terveen, *Using frequency-of-word in public conversation for social filtering*, AT&T Research, 1996. http://weblab.research.att.com/phoaks/phoaks.cscw96.ps

[7] B.A. Huberman et al.: *Beehive: A System for Cooperative Filtering and Sharing of Information*, Xerox Palo Alto Research Center, 1996.
ftp://parcftp.xerox.com/pub/dynamics/beehive.ps

[8] F. Halasz, and M. Schwartz (1994): *The Dexter Hypertext Reference Model*, Communications of the ACM, Vol. 37, No. 2, Feb. 1994, pp. 30-39. (first published in 1990)

[9] H-J. Knappe, *Informations- und Kaufverhalten unter Zeitdruck*, Frankfurt, Bern 1981.

[10] K. Lagus, et al.: *Digital Libraries with the Aid of Self-Organizing Maps*, Helsinki University of Technology, Poster Proceedings 5th Int'l. WWW-Conference 1996, pp. 71-79.

[11] T.W. Malone, et al.: *Intelligent Information-Sharing Systems*, Communications of the ACM (1987) Vol. 30 No. 5, pp. 390-402.

[12] M.Morita et al.: *Information filtering based on user behaviour analysis and best match text retrieval*, Proc. 17th Annual Intl. ACM-SIGIR Conference, Springer-Verlag, July 1994, pp. 272-218.
http://www.jaist.ac.jp/jaist/is/labs/shinoda-lab/papers/1994/sigir-94.ps

[13] D.W. Oard and G. Marchionini: *A Conceptual Framework for Text Filtering*, University of Maryland, Technical Report EE-TR-96-25 CAR-TR-830 CLIS-TR-96-02 CS-TR-3643, May 1996.
http://www.ee.umd.edu/medlab/filter/papers/filter.ps

[14] B. Pras: *Explaining Consumer Decision Making through Evaluation Process Models*. In: Tropitzhofer,E. (Hrsg) 1978, pp. 145-161.

[15] P. Resnick, et al.: *GroupLens: An Open Architecture for Collaborative Filtering of Netnews*, Proceedings of the ACM 1994 Conference on CSCW, Chapel Hill, pp. 175-186.

[16] J. Saarela, M. Turpeinen, T. Puskala, M. Korkea-aho, and R.Sulonen: *Logical structure of a hypermedia newspaper*, Helsinki University of Technology, 1997.

[17] U. Shardanand and P. Maes: *Social Information Filtering: Algorithms for Automating "Word of Mouth"*, MIT Media-Lab, 1995. http://nif.www.media.mit.edu/papers/social.pdf

[18] G. Specht, and M. Bauer (1995): *MultiMED - ein Multimedia-Datenbanksystem zur Aus- und Weiterbildung in der Röntgendiagnostik in der Orthopädie*, in Schoop, Witt, Glowalla (Ed.): Hypermedia in der Aus- und Weiterbildung, UVK-Verlag, Konstanz, pp. 209-210 (in german).

[19] G. Specht: *Complexity Analysis of Link Navigation in Dexter Based Hypermedia Database System*, International Journal INFORMATICA, Vilnius, Vol. 9 No. 2, 1997, pp. 1-20.

[20] C. Stevens: *Knowledge-Based Assistance for Accessing Poorly Structured Information Spaces*. PhD thesis, University of Colorado, Department of Computer Science, Boulder, 1992.
http://www.cs.colorado.edu/home/stevens/public_html/mypapers/Thesis-tech-report.ps

[21] P. Wright: *Consumer Chocie Strategies: Simplifying vs. Optimizing*, Journal of Marketing Research (JMR) Vol. 12 (1975), pp. 60-67.

[22] T.W. Yan and H. Garcia-Mollina. *SIFT — a tool for wide-area information dissemination*, Proceedings of the 1995 USENIX Technical Conference, 1995, pp. 177-186. ftp://db.stanford.edu/pub/yan/1994/sift.ps