D2.7
State of the Art: Personalization

<table>
<thead>
<tr>
<th>Project title:</th>
<th>Knowledge in a Wiki</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project acronym:</td>
<td>KIWI</td>
</tr>
<tr>
<td>Project number:</td>
<td>ICT-2007.4.2-211932</td>
</tr>
<tr>
<td>Project instrument:</td>
<td>EU FP7 Small and Medium-Scale Focused Research Project (STREP)</td>
</tr>
<tr>
<td>Project thematic priority:</td>
<td>Information and Communication Technologies (ICT)</td>
</tr>
<tr>
<td>Document type:</td>
<td>D (deliverable)</td>
</tr>
<tr>
<td>Nature of document:</td>
<td>R (report)</td>
</tr>
<tr>
<td>Dissemination level:</td>
<td>PU (public)</td>
</tr>
<tr>
<td>Document number:</td>
<td>ICT211932/SRFG/D2.7/D/PU/b1</td>
</tr>
<tr>
<td>Responsible editors:</td>
<td>Peter Dolog</td>
</tr>
<tr>
<td>Contributing authors:</td>
<td>Fred Durao, Peter Dolog, Karsten Jahn</td>
</tr>
<tr>
<td>Reviewers:</td>
<td>Norbert Eisinger, Jozef Holy</td>
</tr>
<tr>
<td>Contributing participants:</td>
<td>AAU</td>
</tr>
<tr>
<td>Contributing workpackages:</td>
<td>WP2</td>
</tr>
<tr>
<td>Contractual delivery:</td>
<td>31 August 2008</td>
</tr>
<tr>
<td>Actual delivery:</td>
<td>29 August 2008</td>
</tr>
</tbody>
</table>

**Abstract:**
This is report surveys KIWI-related work in the personalization domain. It discusses various approaches from adaptive hypermedia, user modelling, and semantic web which deal with personalization. It gives some suggestions on potential personalization strategies which can exploit knowledge structures about user and knowledge resources.

**Keyword List:**
Semantic Web, Personalization, User Modelling, Personalization Strategies
# Table of Contents

Table of Contents.................................................................................................................................................. 2  
Foreword: Purpose of this Report.................................................................................................................................. 3  
1. Introduction.......................................................................................................................................................... 4  
2. Personalization in Web Applications ..................................................................................................................... 5  
   2.1. Adaptive Web Based Hypermedia .................................................................................................................. 7  
   2.2. User Modelling ............................................................................................................................................... 8  
   2.3. Semantic Web and Rule Based Personalization ............................................................................................ 10  
   2.4. Evidence from Existing Systems .................................................................................................................. 11  
   2.5. Group Personalization .................................................................................................................................... 15  
   2.5.1. Group Formation in Education Learning Systems ...................................................................................... 16  
   2.5.2. Semantic Group Personalization ........................................................................................................... 19  
   2.6. Statistical Analysis Applied for System Personalization ............................................................................. 21  
3. Prospecting Personalization Activities in KiWi ....................................................................................................... 23  
4. Conclusions ......................................................................................................................................................... 26  
5. References ............................................................................................................................................................ 27
**Foreword: Purpose of this Report**

The Personalization State of the Art document serves several goals:

- Introducing relevant aspects for personalization area in wikis for software development (Section 1).
- Reviewing and understanding the representative work from the area of semantic Web, adaptive hypermedia, user modelling. Discussing relevant considerations about group personalization and outlining the importance of statistical analysis over data for personalization activities (Section 2).
- Giving some preliminary suggestions for strategies which could be used for exploitation of knowledge structures for personalization in KiWi (Sections 3).

This document is intended for technological partners to understand how for example the personalization can be achieved, how it relates to information extraction and reason maintenance and semantic wiki technology. It can also be useful for technologically oriented user partners to see how personalization can be achieved and which scenarios can motivate and demonstrate the personalization.
1. Introduction

With the advent of the Internet, vast amount of knowledge is spread through the Web sites for human consumption. This is also the case of the enterprise Intranets. The growth of the electronically represented information which is shared and prepared collaboratively represents an answer to the needs of companies and individuals starving for just-in-time and up-to-date information. However, although a huge amount of information is available, to find relevant data still remains a difficult task. Possible solutions to overcome this inconvenience are the investment on search and retrieval tools or the creation of sites focused on specific subjects which naturally inhibits users not interested in such content. A more interesting alternative, however, is customization in which the Web sites are employed to provide pertinent content to the right person. In addition of increasing its attractiveness, this solution may result in profitable advantages for the companies behind of the Web sites once they are providing best service to each customer specifically.

A variety of personalization technologies may be applied to achieve the most efficient way of presentation such as adaptive hypermedia, semantic web and user modeling. They have been widely exploited to design adaptive user interfaces in several areas, such as intelligent tutoring systems (ITS), the generation of electronic catalogues, information filtering and recommender systems.

Wiki is a new kind of collaboration, which has widely spread since its introduction to the work group of “design patterns” [20] by Ward Cunningham in 1995. There are many different wikis available these days with many different features and functionality. Even the scopes of them are very different. Wikipedia, for instance, is a very good example for wikis in general, for shared group collaboration or an encyclopedia-style wiki. However, in enterprise activities like software development, Wikipedia is not an example for a wiki that works very well.

A wiki is a simple, Web-based application, that allows every participant (e.g. employees, team members, stakeholders) to create or edit content on pages. This makes it easy to use for collaboration, as different persons can work on the same page, the same content. The wiki meanwhile keeps track of the different versions of pages. So, every change can be viewed and restored at every time.

Software development is a highly collaborative activity that is usually performed by one or more teams, in a longer period of time and in several stages. In every stage, team members need to share knowledge about actions they have performed, solutions to the problems they have introduced, and so on. Traditionally, knowledge in software
development has been shared through documentation. Each result at any stage of software development project was documented and stored either on a desktop or a shared server directory.

Recently, wikis became more popular for the task of knowledge sharing in software development [32]. However, practical experiences by KIWI partners indicate that this approach may lead to a proliferation of wikis, which contain a huge amount of information and make it more and more difficult to find relevant data in them, especially for new team members. Thus, the general problem described at the beginning reappears. Like in the general case, the problem may be overcome by customization or personalization.

Personalization is in general a decision on what might be the most useful for a particular user or a community of collaborating users in specific situations according to what a personalized system observed about a user or a community of users. Therefore, personalization should rely on the contents' and user's semantics. Using specific rules, it is thus possible to present this information to (human) users that fit to their interests and/or roles. The common name for such a data presentation is "personalization", even though "adaptation" to tasks and roles would often be a better name. As we already envisioned in the KIWI vision deliverable D8.5, the semantics of contents and user will bring knowledge necessary to decide on:

- which links should be included or suppressed on the user interface or in the content in particular situations
- which facets of a user interface are relevant for particular situation
- what are suitable configurations of all those

In this report we have tried to review rule-based personalization and user modelling approaches from different application areas which can be relevant for the KIWI project. We scanned number of article in related domains and selected those particularly most relevant to the task of KIWI project. In section 2, an overview about personalization in Web applications is introduced, surveying the representative work from adaptive hypermedia, user modelling and semantic web area related to personalization. In addition special attention is given to group personalization and how statistic analysis may be useful when applied to personalization. Section 3 presents some preliminary suggestions for personalisation strategies which could be used for exploitation of knowledge structures in KIWI system and finally Section 4 concludes the study and raises some reflections.

## 2. Personalization in Web Applications

Adaptation and personalization is the ability of the system to adapt to various
characteristics of a user or user group. Personalization was studied from different perspectives in different communities such as adaptive hypermedia, collaborative filtering information retrieval and so on. However, regarding KiWi Vision Deliverable D8.5, this study relies on personalization of Web applications focussed on building customized hyper documents (web sites) with the aim at meeting the user’s needs more effectively and to increase user satisfaction. Unlike traditional Web applications which simply employ the cookie information to provide personalization, in the KiWi system we can go beyond that and can use additional technologies such as: Web hypermedia, User Modelling, Semantic Web, Group Personalization and Statistical Analysis of user actions.

Web Hypermedia provides the knowledge on how to organize hypertext into interlinked space of information and how to include other kinds of media such as images, audio clips and video clips in addition to the text. User Modelling is an essential activity to maintain information about the user's knowledge, beliefs, goals and plans for achieving his goals, and exploiting abilities, attitudes, and preferences. Personalization decisions can be formally represented by Semantic Web rules mutually with ontology constraints for deciding e.g. what links to show, hide, or recommend based on the most diverse scenarios. Group Personalization will bring knowledge necessary for group formation and individual satisfaction. And finally applying statistical analysis over raw data increases the chances of effective personalization in the system.

Usually Web systems which adopt personalization actions do not carry out whole technologies as seen in Fig. 1. The usual configuration utilizes Web Hypermedia technologies and user models with some business rules applied to browsing content according to user preference. A more advanced configuration however should be achieved adding rule-based personalization through Semantic Web technologies. Statistical analysis
on user history data can be used to discover patterns of user behaviour and to derive knowledge items necessary for reasoning on personalization actions. Furthermore, group personalization techniques should be applied for example for group formation or for raising satisfaction of the group as a whole. An overview of all of these technologies separately with practical evidences will be explored in the following subsection.

2.1. Adaptive Web Based Hypermedia

One limitation of traditional “static” hypermedia applications is that they provide the same page content and the same set of links to all users. In the scenario where the user population is relatively diverse, a traditional system will suffer from an inability to be “all things to all people”. For instance, a static electronic encyclopedia will present the same information and same set of links to related articles to readers with different knowledge and interests. A Web bookstore might also offer the same selection of “bestsellers” to customers with different reading preferences. Finally, a static virtual museum will offer the same “guided tour” and the same narration to visitors with very different goals and background knowledge.

According to [9] review, six kinds of adaptive hypermedia systems were identified: educational hypermedia, on-line information systems, on-line help systems, information retrieval hypermedia, institutional hypermedia, and systems for managing personalized views in information spaces. Until 1996 adaptive hypermedia systems adapted their content exclusively with user characteristics, however, currently the situation has changed. A number of adaptive Web based systems are able to adapt to something else than only user characteristics. [24] suggests to distinguish adaptation to data at three levels:

- **User data:** comprise the traditional adaptation target, various characteristics of the users;
- **Usage data:** comprise data about user interaction with the systems that cannot be resolved to user characteristics (but still can be used to make adaptation decisions);
- **Environment data:** comprise all aspects of the user environment that are not related to the users themselves;

In spite of this conceptual classification which delineates three levels of user characteristics used in adaptation, Web systems usually still face obstacles to discern such differences. Either they do not realize these differences or assume all-in-one.

Perspectives for extending the adaptation techniques of static hypermedia
applications to Web based hypermedia applications are foreseen in terms of integration with other applications and open corpus adaptive hypermedia where this systems in the future should be able to “understand” hyper documents and links to some extent without the help of a human indexer.

A success case of user models for personalized hyperlinks can be seen for example in [6]. The work presents a system that combines adaptive hypertext linking based on group link preferences with an implicit navigation-based mechanism for personalized link recommendations. Basically the system utilizes three learning rules that changes hyperlink weights according to user’s overlapping navigation paths and causes a hypertext system’s link structure to converge to a valid group user model. A spreading activation recommendation system generates navigation path based recommendations for individual users. The system takes into account user’s information need, personal background, and knowledge level and retrieval strategies.

The interesting question of the work is how to aggregate the collective knowledge of a group of users and use it to improve and personalize the experience of individual users? The proposed approach is based on the following methodology:

1. A methodology to transform hyperlink structure into a group user model;
2. Page relatedness measures from simple recordings of user retrieval paths (sequences of retrieved hypertext pages);
3. Guided user navigation by spreading activation recommendations.

The notion of representing a hypertext network as a directed graph is extended with the concept of weighted hyperlinks; each hyperlink in the network is associated with a weight expressing the degree of associative relation between two hypertext pages.

The main contribution of the work is to combine the know-how of the entire group of users stored in the network link structure with the interests of an individual user implicitly expressed in their navigation paths. This whole infrastructure follows a well defined user model which identifies user particularities to represent know-how of the users.

2.2. User Modelling

Certainly user modelling has been regarded as an essential topic of interest for those who desire to implement personalization features in their Web sites. Decomposing user models along multiple aspects help information systems to select the most appropriate news with precise detail level to each individual in particular.

A user model case can be seen in [4] where a framework for the generation of adaptive hypertexts for accessing on-line news servers is presented. The news servers
provide a huge amount of information concerning different subjects. The aim of the proposed system is to present the most appropriate set of news (and advertisements) to each user, choosing the “exact” detail level for each news item. [4] shows that displaying the "right" banners to each user would make advertising more effective, attracting the interest of companies. In a nutshell, the system interaction begins when a new user connects to the server, he fills in a form asking a few initial data to be classified in predefined stereotypes, and the predictions that are generated to constitute the initial user model. If the user has previously registered at the news server, his model is retrieved from the user's database. Given the user model, the system selects the appropriate sections/news, the detail level for the presentation of the news of each section and the advertisements, and dynamically generates the pages. The user's choices during the navigation are recorded; these data may activate the rules for the dynamic revision of the user model and this may in turn change the news and advertisements that will be presented subsequently.

In fact, user models are basis for any system decision and strategies for its building must comprise the entire structure of a Web system from its design up to the presentation layer. Usability goals also are essential aspects that must be carefully discussed and designed in order to avoid elimination of the benefits of adaptivity.

Therefore, [17] presents approaches for dealing with usability challenges frequent in systems based on user modelling and adaptivity. In the work some restrictions are applied because usability problems sometimes outweigh the benefits of adaptation. [17] outlines that preventative anticipation are essential actions of the iterative design of user adaptive systems. This strategy, however, requires a comprehensive understanding of the reasons for typical usability problems and of strategies for preventing them. In a nutshell, the generally desirable usability goals such as “Exact layout and responses” are often threatened by the typical properties of user-adaptive systems such as “Incompleteness of relevant information for adaptation”. Each of the preventive measures such as “Acquire as much relevant information as is feasible” may be able to modify a typical property so as to reduce its negative impact on usability. The compensatory measures such as “Allow inspection of user model”, then, can increase the likelihood that the usability goals are fulfilled even if the threats created by the typical properties cannot be fully prevented.

Preventions and user model strategies are encouraged to be systematically applied in systems which prime for personalization and adaptivity. Not different, semantic web resources, explored in next section, also provide means to build effective user models and reasoning over such structures.
2.3. Semantic Web and Rule Based Personalization

Semantic Web technologies provide formal representation for knowledge on the Web, thereby enabling user modelling as basis for system personalization. Domain ontologies are useful instruments to generate links suited for a particular user interest or simply hiding links not suitable to his goals. Conceptually, domain ontologies consist of classes and relationships used in annotations of specific documents or resources on the Web.

According to [15], user model ontologies provide additional means for deciding which links to show, annotate, hide, generate, and reorder. The semantic web formalization enables reasoning for personalization decisions. Personalized information access in this context is concerned with user-centered bias of the hyperlinks to better support the current user. Furthermore, deduction rules can be employed for personalization and reasoning techniques can then work on metadata based on these ontologies, and generate links based on content, user context and user background.

Technologies to support the maintenance of ontologies has become popular such as Resource Description Framework (RDF) [11] that specifies a simple model for knowledge representation. RDF Schema (RDFS) [2] adds additional expressive power and semantics to this basic model. Ontology Web Language (OWL) ¹ extends RDF properties and metadata besides other things by allowing rules to be part of ontology. In addition standardization initiative for digital libraries metadata such as Dublin Core ² has defined a set of predicates which are used for metadata annotations in the bibliographic domain.

As previously said, reasoning techniques can be applied on the ontology models to infer new facts or some information that has not been explicitly told about. Inferencing optimization techniques have been proposed in the literature with the purpose of reducing costs of implementation and increase performance in real time application. The paper “Inferencing and Truth Maintenance in RDF Schema” [9] have addressed several issues concerning RDF inferencing and truth maintenance, in which contrary to earlier claims in the literature, exhaustive forward chaining is a feasible and even scalable approach for computing the deductive closure of an RDF(S) model. Benefits of this approach are low-cost design and implementation, and very cheap query answering, since this task is reduced to simple lookup without inferencing. In spite of these benefits achieved in [9], it is important to observe that more important than having powerful semantic web resources is the ability of designing effective ontology models with good semantics to enable straightforward reasoning.

¹ http://www.w3.org/TR/owl-features
² http://dublincore.org
An evidence of how using semantics for Web-personalization can be realized in the SEWeP framework [18]. It is based on semantic similarities calculated according to the ontology terms. The dynamic system workflow begins when the Web content is getting annotated automatically with ontology terms. This includes a keyword extraction and a semantic characterization, based on the context. Then there is the C-Logs creation and mining. C-Logs are enhanced Web logs and the result of the correlation of them with the ontology terms. The next step is the semantic document clustering, where semantically close documents are grouped. The last step in this cue is the recommendation engine. The recommendation itself can be of different kind: original, semantic or category based. Original recommendations are for new users or simply based on the last n visits. Semantic recommendations are a bit enhanced, as it takes the semantic proximity into account. The category-based recommendations finally are the same as the semantic recommendations, but at the category level.

Other rule based personalization approaches exist. [13] builds on separating learning resources from sequencing logic and additional models for adaptivity: Adaptivity blocks in the learning object metadata and in various other models like the narrative model, candidate groups, etc. define the kind of adaptivity realizable with the current piece of learning content. The main driving force in these models is the candidate groups that define how to teach a certain learning concept. A rule engine selects the best candidates for each specific user in a given context. An early approach for defining architecture for personalization and adaptivity in the Semantic Web has been proposed in [3]. This approach is characterized by the transfer of ownership of Semantic Web resources to the user, and therefore done on the client side. In AHA! [7], several models like conceptual, navigational, adaptational, teacher and learner models are defined. These models can be mapped onto ontologies / taxonomies, to different schemas describing user profile, and to schemas describing the navigational structure. Adaptation model in AHA! uses a rule based language encoded into XML.

2.4. Evidence from Existing Systems

Evidences of the advantages achieved by personalized systems can be seen over many applications which intend to take into account user needs for efficient content presentation. Most of them however are applied in adjacent areas of knowledge such as learning and education.

The Smart Space for LearningTM (SS4L) [16], for example, is a framework that enables personalized access to distributed heterogeneous knowledge repositories. The main goal is to assist learners to choose an appropriate learning resource or activity,
enabling personalized access to federated learning repositories with a vast number of learning offers. These knowledge repositories hold information on learning resources and people, and have the potential to make a vast number of learning resources available. To achieve the expected personalization benefits, the framework utilizes some techniques such as query rewriting and rule-based reasoning and ranking-based recommendation. Rule-based reasoning techniques are supported by formal ontologies developed based on standard information models for learning domains; ranking-based recommendations are supported through ensuring minimal sets of predicates appearing in query results. The evaluation studies show that the proposed solution enables learners to find relevant learning resources in a distributed environment and through goal-based personalization improves relevancy of results.

According to [21], Web-based learning introduces new educational options to improve teaching and learning practices whereas, in the corporate context, it has the potential to reduce learning delivery costs, create more effective learning environments, accelerate time-to-competency, and increase collaboration. Therefore, consuming and working with the “right” content is quite important in order to create effective learning environments, especially in a cost-driven context such as corporate education and training. Personalization then is applied to support users in making the right choices from an expansive list of options. In this framework, personalization is performed in two ways:

- **Query**: Queries can be rewritten to more specific or more general ones; and
- **Results**: Query results can be differently ranked and annotated.

In the first case, personalization happens before a query is submitted to the learning network. In the second case, personalization happens after query results arrive from the learning network.

An application of Interactive Open Learner Modeling (IOLM) for diagnosing and fostering a learner’s conceptual understanding in a terminological domain is introduced in [11]. The IOLM conceives diagnosis as an interactive process involving both a computer system and a learner that play symmetrical (to a certain extent) roles and construct together the learner model. IOLM also elaborates the notion of interaction, which allows for challenging the robustness of the learners’ knowledge and provoking the users’ active engagement in diagnosis (while learners may happen to browse passively through open learner models). STyLE-OLM – an IOLM system in a terminology domain – was developed to validate the approach, it was developed. The STyLE-OLM is utilized to discuss computational and educational benefits of IOLM in terms of improving the quality of the learner model and engaging the learners in reflective activities. According to [14],
involving learners in situations where they can inspect and discuss their models is a reflective activity which leads learners to think about their knowledge, to articulate, validate, and challenge the robustness of their own domain competence. In order to answer the educational and pedagogic purposes, the STyLE-OLM system had to build an agent environment which constantly consults dialogue strategies – rules that suggest which dialogue should be active in which situation. STyLE-OLM utilizes reasoners to infer potential consequences from interaction with users. It elicits agreements and conflicts in the system and learner’s views about the learner. After each interaction the agreements and conflicts between the commitment stores are extracted and used as a source for updating the learning model.

Another intelligent tutoring system (ITS), COLLECT-UML, is presented in [5]. The system introduces a single and multi-user version of a constraint-based ITS entailed in teaching object oriented design using Unified Modelling Language (UML). Besides technical system and system constraints, it shows the results of the evaluation study performed with students of a Software Engineering course. This might be relevant also in the context of KIWI project as the collaboration in software development projects and externalization of knowledge is very important. According to [5], benefits of collaborative problem-solving depend massively on technologies that provide broad support for student’s communication, thereby encouraging students to verbalise their thinking and working together. In this context, reasoning techniques which handle the domain knowledge able to capture and provide opinions of student discussions are essential for a comprehensive tutorial environment. Furthermore, personalization techniques must be applied to drive specific feedbacks to each user which participates in collaborative environment.

Another application of personalization can be seen in [12]. It proposes mechanism to measure the quality of user contributions, control the overall number of contributions and motivate users to contribute with high-quality resources. The goal of this mechanism is to influence individual contributions by adapting the rewards using a model of the current needs and a model the user’s individual reputation in contributing quality resources. Nevertheless, to measure the quality of each contribution accurately is difficult because quality measures are mostly subjective. Centralized moderation is feasible only for small and narrowly focused communities, where members share similar evaluation criteria. For medium or large online communities (e.g. the ones with more than 100 users), a decentralized moderation for quality measurement is necessary. A feasible solution for that is therefore evaluating the quality of resources or comments through explicit user ratings. However, a challenge in all systems that
rely on decentralized moderation is to ensure that there are enough user ratings. Therefore, different types of rewards must be addressed to different types of contribution. The reward initiative was applied in the Comtella (Fehler! Verweisquelle konnte nicht gefunden werden.), an online class resource-sharing system which was used as a test environment.

Fig. 2 shows the front page of the test system. On this page, the information and features provided for the user are: the user’s contribution levels in the previous week and in the current week; the weights for different forms of contributions; personalized messages for the user and community news. The concerning about quality of the contributions motivated the system to adopt some personalization methods. For instance, personalized messages inform the user of the number of articles he or she should share for the current topic and also remind the user to pay attention to the quality of her contributions and ratings when necessary. In particular, the results achieved in Comtella’s experiment showed that users are encouraged to read articles with more ratings mainly considering that this are supposed to have good quality. Moreover, Comtella’s experiment revealed that personalization was an essential factor to motivate people in contribution with more high quality articles.
In this subsection we explored practical cases where personalization was driven to individual purposes, however, it is common to find groups of people who share same interests and preferences. Group personalization therefore comprises an important subject of study in personalization with special attention to social factors in the Web communities.

## 2.5. Group Personalization

In the KIWI system, most of the personalization actions will be addressed to groups of users according to a domain, department, roles and interests. Therefore, one of the key aspects for KIWI satisfying its member needs is to consider the well-being of the group. In this sense a group model will be useful to represent the common knowledge between group members, modeling how a group interacts and group formation based on individual models.

Group model personalization generally involves consideration of dynamic social factors that are not present in individual models, such as interactions and relationships between users. This sort of knowledge however is difficult to be managed since data is usually unstructured and computationally challenging. Group recommendations, for example, must take into account human factors such as user satisfaction and privacy to be well succeeded. According to [28], a recommender system focussed exclusively on individuals basically prime for maximizing individual satisfaction in which only the items with the highest ratings are recommended. In this case, there is no need to predict user satisfaction accurately. However, if we are interested in keeping a group satisfied, accurate prediction of individual satisfaction becomes crucial.

To believe that group recommendations will be always efficient is a utopia because while most members of the group are pleased with recommendation, a single individual might be confronted occasionally with items they are not interested in. However, efforts are made in order to avoid eventual inconveniences. An accurate prediction of satisfaction would help to ensure no individual gets excessively dissatisfied by presenting disapproved items. Actually, prediction is not an exact science with stable expected results; it depends on continuous monitoring of user’s activities and statistical analysis of their preferences. Furthermore, these efforts enable the evaluation group aggregation strategies under various conditions without the need for real users.

In order to measure the user satisfaction, [27] proposed a function based on the summation of satisfaction on preceding items and the impact of a new item based on its rating. Nevertheless, such measurement becomes additionally complex when dealing with groups, because of members of the group influencing each other’s emotions, via emotional contagion and conformity [28]. If in one side single rating statistics may be regarded as
insufficient for realistic recommendation, on the other side, *modelling affective state* is high cost feature to be implemented by any recommender system considering the effort to gather and process such unstructured data. *Privacy* is another obstacle to be overcome by the group recommendation. Group recommender systems may also reveal *private* or *intimate* interests of a specific user to the rest of the group [28]. This has philological impacts given that some disclosures may let users embarrassed and as a consequence let them to avoid additional ratings or even quit the system evermore besides propagate its disappointment to other users. A feasible way to avoid this kind of problem is to assess the level of detail when divulging preferences of the individuals in a group. In fact, there is no best solution to model the state; it depends on how the system cares about its users and which “feelings” must be comprised in the affective model.

In the KiWi, groups will be always frequent considering the fact that software development environments are often creating teams to tackle software projects. Group formation then remains as challenging feature to be analyzed so that good group arrangements can result in more productivity teams. Although not reporting cases directly linked to software development, we present some experiences in the education domain with the aim at demonstrating the way in which students are grouped may affect the results of collaborative work.

### 2.5.1. Group Formation in Education Learning Systems

With the aim of obtaining information about the impact of learning styles on the success of collaborative work, [2] have designed a case study in which the main goal is to gather information about whether the learning styles of several students working together may influence the outcome of their collaborative work and, if so, which are the relevant features that may affect the success of the learning experience. The study adopted the Felder and Silverman model [19] which categorizes a student’s preferred learning style along a sliding scale of five dimensions: sensing-intuitive (how information is perceived), visual-verbal (how information is presented), inductive-deductive (how information is organized), active-reflective (how information is processed) and sequential-global (how the information is understood). The study was carried on with 166 students from a course on Theory of Computation which answered specific questions with the purpose of finding out how the proposed *learning styles* could influence in collaborative learning. This worth information therefore can be used with *adaptation purposes in educational hypermedia and collaborative systems*. The study concluded that some dimensions of the learning style model, namely *active-reflective and sensing-intuitive*, seem to affect the quality of
the resulting work. The outcomes motivated the improvement of TANGOW, a collaborative learning Web-system that supports the creation and delivery of adaptive Web-based hypermedia and dynamic group formation. TANGOW then was improved with new grouping rules with the aim of maximizing the possibility of their success during the realization of collaboration tasks.

Similarly to TANGOW, COOPER (Collaborative Oral and written language adaPtive Production EnviRonment) is learning framework which combines individual and collaborative learning in order to facilitate the second language learning, by enabling students to work individually and then participate in authentic collaborative communicative activities where their linguistic competences can develop, as they do in real ‘immersion experiences’. COOPER utilizes a student and group model required both to structure and to coordinate the way in which the students work together [31].

<table>
<thead>
<tr>
<th>STUDENT MODEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name:</td>
</tr>
<tr>
<td>Native language:</td>
</tr>
<tr>
<td>Type of English:</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INDIVIDUAL ACTIVITIES:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Material studied:</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Initiative:</td>
</tr>
<tr>
<td>Objectives:</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Degree of theoretical explanation:</td>
</tr>
<tr>
<td>Other preferences and/or restrictions:</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>COLLABORATIVE ACTIVITIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>References to group models in which the student has/is working.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INDIVIDUAL L2 LEARNING CHARACTERISATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conceptual units:</td>
</tr>
<tr>
<td>Knowledge stage:</td>
</tr>
<tr>
<td>Linguistic level:</td>
</tr>
<tr>
<td>Linguistic skills:</td>
</tr>
<tr>
<td>Prototypical ling. activities:</td>
</tr>
<tr>
<td>Learning phase:</td>
</tr>
</tbody>
</table>

Fig. 3 A summary of the information represented in COPPER’s student model

The student model (Fig. 3) represents the state of second language learning together with the profile and history of a student’s individual and collaborative activities. It contains the personal details and linguistic background of the student, the individual learning activities undertaken with the system and the collaborative activities undertaken by a student.
The group model (Fig. 4) stores collaborative data and encapsulates the adaptive part of the system, since the group generation is dynamic, based upon a variety of factors, and represents the details of the sets of students working on particular tasks, detailing interactions, mistakes, etc. This model works as a single register that characterizes the collaborative activity of the group. The information contained in this model is useful for two reasons. Firstly, the adaptive group formation considers the previous group activities of each student when forming new groups. Secondly, the logs of these group models will permit the learning process undertaken by each student to be analysed in the future, and will also enable the overall pedagogic properties of the system to be evaluated.

The group formation process is adaptive in the sense that the algorithm is used to dynamically match the students available at a particular moment in the virtual community (using the data in each individual student model), with collaborative templates that define group activities, in terms of the tasks that make up an activity, the roles to be assigned to the students, and the criteria by which the monitor can evaluate both the collaboration process and the results produced. As seen, learning environments have contributed with further information about group formation taking into account the impact of learning styles on the success of collaborative work. In a software development scenario, the knowledge gained by studying the learning systems should be utilized concerning groups formation when creating teams for running software projects. In the KiWi, information
about user competences and experiences should be used to suggest the creation of groups depending on the nature of the project. This activity therefore can save time from project managers when seeking qualified people for being part of the project.

2.5.2. Semantic Group Personalization

Group modelling may be formally expressed by domain ontologies able to draw group relationship properties. In this content, [29] propose a semantic learner model based on the Friend of Friend (FOAF) Ontology\(^3\), a vocabulary for mapping social networks. They believe that group formation must be viewed as a personalization of the individual’s allocation to a group based on the premise that individuals tend to be different and the problem of satisfying all the user’s needs will always exist. Negotiating the allocations to reach consensus is a challenging task considering the fact that each individual has its own experience and preferences. This concerning matches our expectations about group formation in a software development scenario as discussed in the last subsection.

In this study, three types of groups are considered according to its duration, cohesion, and purpose.

<table>
<thead>
<tr>
<th>Types of Groups</th>
<th>Meanings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team</td>
<td>Planned group with oriented tasks with short duration.</td>
</tr>
<tr>
<td>Communities</td>
<td>Usually represented by groups of people which have a common interest in some subject.</td>
</tr>
<tr>
<td>Networks</td>
<td>Collaborate to share ideas or find solutions and Networks in which the members in this group do not need to be familiar with each other as long as they can collaborate to deliver the task.</td>
</tr>
</tbody>
</table>

To each type of group shown in Table 2, [29] propose a Semantic Learner Ontology (SLO) with task-related parameters to model the students in relation to the course or the task of the group work such as experience, education level, knowledge, skills, abilities (cognitive and physical), grades, interests, and preferences of topics; relation-oriented parameters with personal information on the student such as gender, age, culture, social status, personality; and context-related parameters which involve information on the context features of the students and their environment such as geographical location, availability schedules, and communication tools.

\(^3\) http://xmlns.com/foaf/spec
Fig. 5 The Semantic Learning Ontology

Fig. 5 shows the FOAF ontology describing its classes on the right side and the respective properties on the left side. The proposed ontology may serve as inspiration for adaption (and extension) to other domains including software development. According the KiWi needs, the proposed group categorization can be inherited from the learner’s models to a software development organization due to the fact that software projects commonly are tackled by distributed teams geographically separated. Additionally, open source projects usually depend on external contribution where the members do not need to be familiar with each other as long as they can collaborate to deliver the task.

[34] also propose an ontology in the domain of software engineering with the aim at enhancing software development project information by associating the project information with specific entities within the domain of interest. The software engineering ontology organises and centralises software engineering knowledge in a formal, machine and human understandable way whereby software applications or a software agent can use the knowledge regarding project information as instances in the ontology to carry out knowledge maintenance.
According to the Fig. 6, the specific software engineering concepts used for the particular software development project are captured in the specific software engineering ontology as sub domain knowledge. The generic software engineering ontology represents all software engineering concepts while specific software engineering ontology represents some concepts of software engineering that the particular project needs. Then in each project there exists project information or actual data including project understanding and project agreement.

In the KiWi, Semantic Web applications should match the project requirements with software engineering concepts held in the ontology to suggest similar project development data from previous experiences.

2.6. Statistical Analysis Applied for System Personalization

The effectiveness of personalization in a system depends on the quality of data that is being utilized. Therefore, Web systems which rely on statistical analysis over raw data from user activity may increase the chances of successful personalization. The interpretation of this data addresses the system to efficiently e.g. identify clusters of users based on behaviour and ignore malicious visitors which do not contribute at all. In addition, it is easier to identify hidden group patterns and new trends of user behaviour that occasionally may appear influenced by e.g. new technological fashions or periodical events.

According to [26], understanding user behaviour and discovering the valuable information within vast amount of information involves several phases: data structuring and storing, data cleaning and pre-processing, where typically noise is removed, log files are broken into sessions and users are identified; data transformation, where useful features are selected to represent the data and applying data mining techniques to identify interesting patterns, statistical or predictive models or correlations among parts of data.
Finally, the interpretation of the results becomes a very precious activity because implicit knowledge is discovered and then transformed in requirements for future personalization implementations. To find out the communal users behaviour is important for a number of tasks such as managing the site, targeted advertising, and identifying malicious visitors. It also helps understanding the navigation patterns of different user groups and therefore helps in organizing the site to better suit the users.

Motivated by some benefits discussed previously, [26] investigates the use of maximum entropy mixture models and mixture of Markov models for inferring individualized behaviour models of Web users. The behaviour model is here represented by a probabilistic model describing which actions the user will perform in the future e.g. read an article, rate content, or edit its profile. For the evaluation, a dataset consisted of CiteSeer log files covering a period of approximately two months containing information about user’s actions was analyzed. The interpretation of clusters obtained in the experiments concluded that maximum entropy and Markov mixture models have both descriptive and predictive power for representing user behaviours, and a useful mechanism for interpreting user behaviour patterns for the Web data. In particular, they were able to identify and visualize specific behaviour patterns of CiteSeer users and argue that maximum entropy model’s computational cost recompense at recognizing complex dominant patterns of user behaviour.

Starting from the analysis of limits of group modelling strategies, [30] presents an experience of building a system to give news adapted to different group of users in a public space. They suggest an update of the probabilistic group model to improve the interaction of groups of users with devices devoted to show news. In the evaluation test of the system, they wanted to investigate whether contextualizing the interaction of a public kiosk was effective or not. To efficiently personalize the exhibition of news, a group division was carried out by statistical information gained using a questionnaire and the self selection undertaken by one algorithm of group modelling. The experiment has demonstrated that the interaction using a group modelling adaptation is promising, because users involved have shown more satisfaction in watching filtered news, rather than not filtered. Moreover, the study revealed that during the interaction they were motivated to argue each other about the topics shown into the display. This system shows that integration between personal and group integration is possible, and users like to be socially involved in discussion about common interests. Finally, by analyzing the questionnaire compiled by the users which participated of the evaluation, [30] outlines that usually in social

---

4 http://citeseer.ist.psu.edu
interaction there is always someone, or a small group, that influence the other majority. People interacting in a personal mode make different choice than in group. So the group takes decision using the Most Respected Person Strategy, delegating to this minority the choice.

Lekakos and Giaglis [25] propose recommendation approaches that follow the collaborative filtering reasoning and utilize the notion of lifestyle as an effective user characteristic that can group consumers in terms of their behaviour. Starting from a basic lifestyle-based recommendation algorithm, [25] incrementally develops hybrid recommendation approaches that address certain dimensions of the sparsity problem and empirically evaluate them providing further evidence of their effectiveness. The lifestyle approach represents the fundamental construct toward the improvement of current personalization approaches. However, the most important contribution of the lifestyle approach is that it works as the key constituent of the subsequent proposed approaches which further improve the predictive accuracy of collaborative filtering algorithms in sparsity conditions. Furthermore, it can accommodate any type of personalization approach that operates upon ratings and improve its performance through the reduction of the sparsity effect.

Significant studies involving statistical analysis have been taken aiming at improving the personalization efficiency. Diverse techniques and algorithms have been applied in Web systems to provide more accurate data used for personalization purposes. In the KiWi, e.g. Neural Network algorithms could be applied on user’s navigation partners in order to classify user profile according to technical interests and then organizing the link structure to better suit them. Moreover, a statistical analysis over the user’s log could be used for identifying the most and the least active users and sensibly send them personalized messages of congratulation and motivation respectively.

3. Prospecting Personalization Activities in KiWi

In the KiWi, personalization activities will be fundamental for effective knowledge management and collaborative work. A range of the techniques introduced in this study such as Collaborative Filtering, Group Recommendation and Semantic Web reasoning will be reasonably analysed and explored with the aim at providing information tailored according to the user’s preferences and needs.

Web Hypermedia. Most of the personalization activities will affect KiWi interface. In
this sense, hypermedia knowledge is necessary to organize the KiWi Web pages in terms of link structuring and media allocation. Regarding the software development scenario where users differ in terms of roles, a project manager will likely find project schedules and spreadsheets about the current resources in his home page whereas the programmer will likely be addressed to pages containing technical issues or bug information.

**User Modelling.** KiWi in its essence is a user-centred system; therefore the KiWi´s user model must be able to maintain information about the user's knowledge, beliefs, goals and plans for achieving these goals, abilities, attitudes, and preferences. Decomposing the user model along multiple aspects will help KiWi to select the most appropriate piece of information with precise detail level to each individual in particular.

**Statistical Analysis.** Starting from the point of gathering relevant data, KiWi must be constantly monitoring, storing and then analyzing user data with the purpose of identifying groups of users according to their preferences and behaviour. Furthermore, we understand that statistical analysis on general Wiki data is a crucial activity for discovering more valuable information within vast amount of information in the knowledge space. Specific algorithms for data classification and knowledge engineering such as Neural Network or Naive Bayes will allow the discovering of information partners which become inputs for personalization purposes such as navigating, targeted advertising, or identifying malicious behaviours. It also may be utilized to capture navigation patterns of different user groups and therefore helps in organizing the site to better suit the users.

**Group Personalization.** After having data processed, personalization actions are prospected to be better succeeded than when managed using unrefined raw data. Then, KiWi should perform group recommendations taking into account knowledge about resources and user experience including here collaborative filtering statistics based on user rating. This action will require however a comprehensive user model which satisfies strategies to compose the user groups and their relationships. Moreover, an accurate prediction of satisfaction about the recommendation will be needed to avoid displeasure within the user community. Motivation mechanisms also are expected in order to incentive individual contributions by rewarding them based on their current needs. Thus, we prospect to have personalized messages of incentive, awards for the most active contributors and rating of users.
**Semantic Web and rule-based Personalization.** Semantics appear as a fundamental requirement in the KiWi, considering the fact that modelling the published data and the user profile with ontologies allows expressing more effectively the user interests and the relations between the pieces of information. Furthermore, some of the semantic relations may be exploited for the creation of rules applied to new personalization decision such as enabling the generation of links suited for particular user interest or simply hiding links not suitable to his goals. In addition, the semantics of information e.g. will bring knowledge necessary to decide on which links should be included or suppressed on the user interface and which facets of a user interface are relevant for particular situation. In the following subsection we propose a preliminary recommendation process which regards knowledge about resources and user models.

**Exploitation of Semantic Structures for Personalization.** Similarly to SS4L [16], semantic graph structure of user, group, and knowledge resources (pages) can be exploited.

![Fig. 7 Personalization in SS4L [16]](image)

Fig. 7 exemplifies the recommendation process which is performed as a matching function between knowledge about resources and knowledge about a user. The left side of the figure depicts the conceptual model for Alice’s learning performance in terms of programming in Java and .NET courses. These courses extended Alice’s knowledge. The right side of the figure depicts a learning resource on advanced security with a subject referring to the selected learning goal in the query. The middle part of the figure refers to a
partial schematic view on heuristics which have been selected to annotate the learning resources. In this example, all prerequisites of the activity match Alice's performance record; therefore the learning activity will be annotated as recommended.

In the proposed learning-performance-based recommendation, the semantic graph of learning resources returned as query results is examined and if there exist a matching between prerequisite relations in the learning resource and performance relation in learner profile, the resource is recommended. The relationships from both sides can be explored transitively. In the KIWI system, we need to study the properties and search strategies over the space of knowledge resources (pages) and users also grouped into groups according to a domain, department, roles, and interests. Another possibility is to study strategies on which subset of personalization rules apply in which context. [17] suggests constraining the personalization rules by domain and user preferences. The personalization rules are applied to a query rewriting. The query rewriting/personalization rules are annotated with preference patterns or queries over preference graphs and query patterns which trigger them. This is also relevant to the KIWI project, however, the model for personalization need to be redefined to the knowledge resources and user actions on the user interface instead of queries.

4. Conclusions

Ever more the World Wide Web is being populated with numerous online applications. In these applications a user may interact with a service provider, product sellers, governmental organizations, friends and colleagues. A vast amount of content and services are available at different sources and places providing different types of information and knowledge to user. This multiplicity of available information however needs to be reasonably addressed to personalized consumption. Personalization and adaptation techniques allow Web applications to combine all available knowledge in order to form personalized, user-friendly, and business-optimal services.

Personalization strategies allow users to filter large amounts of information implicitly besides reducing information retrieval usage. In addition, it provides the compelling user experience to increased site traffic, participation and value. Economically, personalization techniques may be addressed to those sites whose products are expected to achieve considerable gains. Moreover, personalized portals can have their return on investment (ROI) maximized by saving content administration by reducing administration operations costs while increasing content exposure and their associated revenues.
A number of technologies behind personalization may be set up to make it more efficient. **Statistical analysis** on user data can provide refined information or help the system to discover user’s behaviour patterns. **Semantic Web** resources through domain ontologies are able to formally represent **user models** and maintain personalization decisions by explicating **personalized rules**. However, continuous maintenance is needed since users scenarios are constantly changing. **Group personalization** techniques may be explored concerning group formation regarding the fact that users in a group should at some point share common interests. Privacy remains an open issue to be more explored since information in a group may reveal confidential preferences to everyone. All of these technologies provide worth contribution to better answer the user’s needs and make personalization more reliable.

5. **References**


Intelligent Information Systems. 2008


(Malta, September 15 - 17, 2005).


