Integrated Project

Priority 2.4.7

Semantic based knowledge systems

Information Society Technologies

The Social Semantic Desktop

NEPOMUK

Mandriva Community Scenario Report

Deliverable D11.1

Version 1.2

01.02.2007

Dissemination level: PU

Nature	R
Due date	30.09.2006
Lead contractor	EDGE-IT S.A.R.L
Start date of project	01.01.2006
Duration	36 months

Authors

Stéphane Laurière, EDGE-IT S.A.R.L
Alexandre Solleiro, EDGE-IT S.A.R.L
Sebastian Trüg, EDGE-IT S.A.R.L
Cristian Bogdan, KTH
Kristina Groth, KTH
Pär Lannerö, KTH

Mentors

Vasilios Darlagiannis, EPFL
Manfred Hauswirth, EPFL
Markus Junker, IRION MANAGEMENT CONSULTING GMBH
Dr. Thomas Roth-Berghofer, DFKI
Gerd Stumme, University of Kassel
Markus Junker, IRION MANAGEMENT CONSULTING GMBH

Project Co-ordinator

Dr. Ansgar Bernardi
German Research Center for Artificial Intelligence (DFKI) GmbH
Erwin-Schroedinger-Strasse (Building 57)
D 67663 Kaiserslautern
Germany
Email: bernardi@dfki.uni-kl.de, phone: +49 631 205 3582, fax: +49 631 205 4910

Partners

DEUTSCHES FORSCHUNGSZENTRUM FUER KUENSTLICHE INTELLIGENZ GMBH
IBM IRELAND PRODUCT DISTRIBUTION LIMITED
SAP AG
HEWLETT PACKARD GALWAY LTD
THALES S.A.
PRC GROUP - THE MANAGEMENT HOUSE S.A.
EDGE-IT S.A.R.L
COGNIUM SYSTEMS S.A.
NATIONAL UNIVERSITY OF IRELAND, GALWAY
ECOLE POLYTECHNIQUE FEDERALE DE LAUSANNE
FORSCHUNGSZENTRUM INFORMATIK AN DER UNIVERSITAET KARLSRUHE
UNIVERSITAET HANNOVER
INSTITUTE OF COMMUNICATION AND COMPUTER SYSTEMS
KUNGLIGA TEKNISKA HOEGSKOLAN
UNIVERSITA DELLA SVIZZERA ITALIANA
IRION MANAGEMENT CONSULTING GMBH

Versions

Version	Date	Reason
0.1	01.09.2006	First draft
0.2	6.09.2006	Personas and new scenarios have been added
0.3	12.09.2006	User observations have been added
0.4	20.09.2006	The ontology specification has been added
0.5	24.09.2006	The conclusion has been added
0.9	27.09.2006	Mentor remarks have been taken into account
1.0	28.09.2006	Released as final
1.1	13.10.2006	Adaptation to new template and final streamlining by IMC, submitted to EC
1.2	01.02.2007	Changed dissemination status to public in agreement with all partners

Explanations of abbreviations on front page

Nature
R: Report
P: Prototype
R/P: Report and Prototype
O: Other

Dissemination level
PU: Public
PP: Restricted to other FP6 participants
RE: Restricted to specified group
CO: Confidential, only for NEPOMUK partners

Executive Summary

The objective of WP11000 is the adoption, application and validation of the services provided by the Nepomuk platform in the context of the large on-line community of Mandriva Linux users.

This report presents a state of the art, a business case, a series of user requirements and of assessment criteria related to the case study.

The reported work has been guided by the following methodology:

- the end users have been directly involved in the process of defining the scenarios and the requirements of the target system,
- a coordination task force bringing together all the case studies has made sure the findings are presented in a coherent manner,
- although the study focuses on the case of the Mandriva Linux community, the findings and the requirements remain generic enough for a potential reuse by other communities with similar needs.

The work is presented in the following chapters:

Chapter 1 – "Introduction" – reminds the context and the objectives of the case study.

Chapter 2 – "State of the art on social semantic help desking" -– studies the adoption of the social Semantic Web vision and of social networking approaches through existing platforms devoted to knowledge sharing between users and experts.

Chapter 3 – "Business case for knowledge management improvement within the Mandriva Community" – outlines the business argument for adoption of Nepomuk technology by the Mandriva community and presents the adoption process. It also presents the new business activities Mandriva can derive from the system.

Chapter 4 – "User requirements" – presents findings resulting from interaction with the end users. It introduces personas representing the Mandriva Linux users, and defines a series of concrete user requirements derived from scenarios presenting the help desk at work.

Chapter 5 – "Case study domain ontologies" – presents the approach to work that will be adopted for drafting the domain ontologies. Several existing domain ontologies related to the case study are presented.

Chapter 6 – "Assessment Criteria" – describes the methodology for measuring the evolution of the case study both quantitatively and qualitatively.

A key result derived from the literature review and from the Semantic Web initiatives conducted in the realm of life sciences, is that, to lower the drafting cost of the large domain ontologies required by the case study and to maximize their impact, the ontologies will be designed

collectively using a semantic wiki favouring an open, parallel, decentralized, and synergistic protocol.

The report also emphasizes the need of a platform for sharing knowledge on free software not specifically related to a given Linux distribution: such a platform is likely to speed up the process producing a large amount of multi-lingual material organized in a semantic manner. It is envisaged to team up with the SELF EU project on this matter.[1]

The study of existing knowledge sharing platforms has shown that one of the flaws of the current Mandriva help desk system is its lack of support for group forming. In providing the infrastructure for creating groups, Nepomuk services will let users take advantage of new paradigms for knowledge sharing and discovery such as "social search".

In collaboration with all other Nepomuk workpackages, WP11000 aims at delivering one of the "killer applications" of the Social Semantic Web.

As the knowledge sharing process is at the core of the production and innovation momentum of the Mandriva community, the effects of Nepomuk achievements in the context of WP11000 are potentially huge. WP11000 will ease dramatically the way end users can get precise answers to their questions, the help desk is likely not only to support the processes of the Mandriva Linux community, but also to have a tremendous impact on the adoption of Linux based systems on the desktop.

---

[1] SELF project web site: http://selfproject.eu/

Table of contents

1. Introduction

The objective of WP11000 is the adoption, application and validation of the services provided by the Nepomuk platform in the context of the large and active open-source on-line community of Mandriva Linux users. WP11000 aims at equipping the members of this community with a new generation tool for sharing knowledge related to the open-source Mandriva Linux project.

The Mandriva Club is an on-line community of approximately 15,000 members. As of 2006, this community can communicate over the main Club web site (http://club.mandriva.com) including the Club forum, which is a minimal help desk system. The main Club web site is entirely wiki based allowing community members to participate easily in the publication of new content and in the maintenance of the whole content set. As for the forum, it consists of an on-line bulletin board with dedicated sections for several languages (English, French, German, Portuguese, Spanish, Dutch and Norwegian).

The current communication system provided by the Club infrastructure is used daily by hundreds of users for finding answers to their questions or for bringing relevant and up-to-date information related to the use of Mandriva Linux operating system, or more broadly related to the prospects of the Mandriva Linux project and ecosystem. This communication system aiming at knowledge sharing and creation can however be considerably improved. WP11000 will build upon the services provided by Nepomuk technical workpackages for creating a tailored environment for knowledge sharing within the Mandriva Club community.

Since the core feature of the system focuses on helping users solve problems they may encounter while using their Mandriva Linux operating system or any software running upon it, the system is a help desk. In the context of Nepomuk, the system is dubbed a "Social Semantic Help desk":

- The help desk will be semantic in the sense that it will take advantage of several ontologies for structuring the contents written by community members. This text structure will not only help the users spell out questions or answers in an unambiguous manner, but will also make it possible to use an advanced search engine across the contents and to reduce the conversion of question/answers threads into reusable knowledge.
- The help desk will be social in the sense that the questions and answers will be brought by the entire community, and the underlying system will provide support for establishing a trust network across the participants.
- The help desk will be tightly integrated with the Mandriva Linux desktop, based on the KDE graphical environment, and with the technological frameworks Eclipse and Mozilla Firefox. This will allow for a rich browsing and editing of user experience, real-time communication, and local annotations of public contents.

2. State of the art on "social semantic help desking"

The scope of this section is the description of methodologies, help desk platforms and communities that may be useful for creating the social semantic help desk targeted by WP11000. The focus is brought to systems that allow the involvement of large scale open communities.

The objective of this state of the art is twofold: first it aims at identifying existing components, APIs or approaches that can either be directly reused for building the final system or serve as a fruitful inspiration source. Second, it aims at identifying the key factors of success and the assessment criteria pertaining to the goals to be achieved.

2.1. Methodology

The adopted methodology consists in selecting most representative platforms among existing ones and to analyse these tools along the core Nepomuk axis. For each core axis, a state of the art is conducted. To the best of our knowledge, there is no such thing as a "Social Semantic Help Desk" on the market yet. Several frameworks and platforms however address separately each of the three cornerstones of the target system: the social one, the semantic one, or the one that purely relates to help desking. The most relevant platforms are presented below.

Core axis	Description
Semantic web and knowledge management axis	This axis relates to the support of semantic contents (tagging and ontology support), search tools, mechanism for converting questions/answers into reusable knowledge, availability of a contextual knowledge base, text analysis.
Social axis	This axis mainly relates to recommendation systems, trust frameworks, identity management, social search and discovery, capacity to form groups.
Desktop integration axis	Desktop integration relates to the possibility of using the software service from rich user applications such as desktop office tools, rich email clients, rich calendar and task management clients.
Openness and extensibility axis	This axis relates to the openness and extensibility of the platforms in terms of source code access policy and availability of open APIs accessible over the web.

Table 1: Core axis of interest

2.2. Most representative help desk platforms

The table below presents an overview of representative help desk platforms and products in the context of the study.

Platform / Product	Short description
Microsoft QnA http://qna.live.com/	Windows Live QnA is a question and answer service, part of Microsoft's Windows Live range of services. It is somewhat similar to Google Answers, but closer in function to Yahoo! Answers. The service is currently in a public beta, and is available only to US residents for testing. Users can ask questions, tag them according to topic, and gain points and reputation for answering other users' questions. (source: Wikipedia)

Platform / Product	Short description
Google Answers http://answers.google.com	Google Answers is an Internet search and research service offered for a fee by Google. It is an extension to the conventional search: rather than doing the search themselves, users pay someone else to do the search. Customers ask questions, offer a price for an answer, and researchers answer them. Researchers are not Google employees. They are limited in number (according to Google, there are more than 500 Researchers; in practice, there are probably fewer active Researchers) and are screened through an application process that tests their research and communications abilities. Prices for questions range from $2 to $200; After a question is answered, Google keeps 25% of the payment, and sends the rest to the Researchers. In addition to the Researcher's fees, a client who's satisfied with the answer can also leave a tip. (Source: Wikipedia)
Experts-Exchange http://www.experts-exchange.com/	Experts-Exchange, founded in 1996, is a collaboration platform for information technology professionals, designed to address specific areas of situation-based knowledge. It's a fee-based online "ask an expert" site for computer related questions, with limited free access also available. (Source: Wikipedia)
Qunu http://www.qunu.com	Qunu is a help-desk system based on instant messaging. It lets search for experts available online for real-time help. As states on Qunu's web site: "Instead of being left with more questions than answers after trawling through search engine results, Qunu provides a direct line to those who not only know their stuff, but who are also passionate about sharing it - with you!"
Yahoo! Answers http://answers.yahoo.com/	Yahoo! Answers is a community-driven service that allows users to ask and answer questions posed by other users.
SIPHS "A Life Science Community" http://www.siphs.com	"The SIPHS mission is to expedite scientific discovery by facilitating information sharing and open dialogue within the biological and biomedical sciences community. It's a simple system really: 1. Search for answers to your questions. 2. Ask your question if you can't find an answer. 3. Share your knowledge with the rest of the community!"
Wink http://www.wink.com	"Wink is a social search engine. We surface the search results that other people have found to be the best stuff for your questions. We have lots of simple ways to let you say what you think about the quality of results - from being able to re-rank search results, to bookmarking sites you want to remember and blocking results that aren't good. Wink's PeopleRank (tm) technology takes all this information and refines it to deliver the results that people think are the best."
Microsoft Support http://support.microsoft.com/	Support.microsoft.com web site is the official professional support platform for end users.
Knova http://www.kanisa.com/products/industry-helpdesk.html	"KNOVA is a leading provider of service resolution management (SRM) applications that reduce service and support costs, increase revenues, and improve customer satisfaction." Built on a next-generation search and knowledge management platform, the KNOVA suite of applications automates the resolution process across multiple channels including contact centers, help desks, email, and self-service sites.
Remedy BMC Remedy	BMC Remedy Customer Support is one of the key industry provider in the field of help-desk software. It is built on the BMC Remedy Action Request System, "a service process management platform noted in the industry for its easy adaptability" (Source: BMC Remedy Corporate Datasheet)

Platform / Product	Short description
Best Pratical Request Tracker http://www.bestpractical.com/rt	RT is an enterprise-grade ticketing system which enables a group of people to intelligently and efficiently manage tasks, issues, and requests submitted by a community of users. The RT platform has been under development since 1996, and is used by systems administrators, customer support staffs, IT managers, developers and marketing departments at thousands of sites around the world.
OTRS http://otrs.org/ http://otrs.org/demo/	OTRS is an Open source Ticket Request System.
Kayako – eSupport http://www.kayako.com/	"Kayako SupportSuite offers integrated Multi-Channel solution allowing you to manage your emails, online issues, chats, self service and issues received by phone. Some key features of the product include: AJAX based Rich User Interface, Active Directory (LDAP), vBulletin, ModernBill integration, ViewShare feature allowing you to guide your clients in realtime, Staff to Staff Voice Chats, SLA's, Work Schedules and Escalations, Detailed Reports and Analytics, Microsoft® Outlook® Integration, Teamwork module allowing you to create shared Events, Contacts and Tasks" (Source: Kayako web site)

Table 2: Short descriptions of most representative help desk platforms and products

2.3. Knowledge management and semantic web axis

2.3.1. Classification schemes

All the major help desk platforms have a classification scheme for organizing the questions and answers. However, as it appears in the table below, the classification schemes used by existing help desks lies globally at the lower end of ontology expressiveness. To the best of our knowledge, there is currently no large community platform on the web that builds upon a domain ontology for representing the submitted questions and answers.

Platform	Classification scheme
Yahoo! Answers	Yahoo! Answers uses categories for classifying the contents. The hierarchy contains 25 top level categories covering all most general topics, approximately 10 second level categories per top level category, and approximately 2 third level categories per each second level category. The hierarchy has then approximately 500 terms. Questions can belong only to one category.
Google Answers	Google answers uses a two-level hierarchy of categories, with 10 top level categories, and approximately 10 sub-categories for each. The Google Answers categorization hence contains approximately 100 terms. Questions can belong only to one category.
Microsoft QnA	Microsoft QnA use tags for classifying the questions. New tags can be entered by the members. The system does not provide information on related tags.
SIPHS	SIPHS uses a large number of tags for classifying the experts. SIPHS does not provide information about related tags. Submitted questions can only apply for one

Platform	Classification scheme
	tag of experts.
Qunu	Qunu uses a large number of tags for classifying the experts. New tags can be entered by the members. Qunu suggests related tags while users search for help.
Microsoft Support	Microsoft support platform classifies knowledge base articles per product, per language, per keyword, and per type of resource (resource types include: how to articles, troubleshooting, guided help, MSDN Articles and TechNet Articles). Each knowledge base article typically contains the list of products and versions it applies to and a list of keywords. Each type of resource also has a dedicated structure. A troubleshooting article for instance will have the following fields: symptom description, cause description, status of the issue, products it applies to and keywords.

Table 3: Description of the classification schemes used by existing help desk platforms or products

As a consequence of low level classification, the available search engines only let users enter high level search criteria, as for instance in the search dialogue below from Microsoft Support web site. Although this type of search is powerful, it has no possibility yet to harness the expressiveness of domain ontologies.

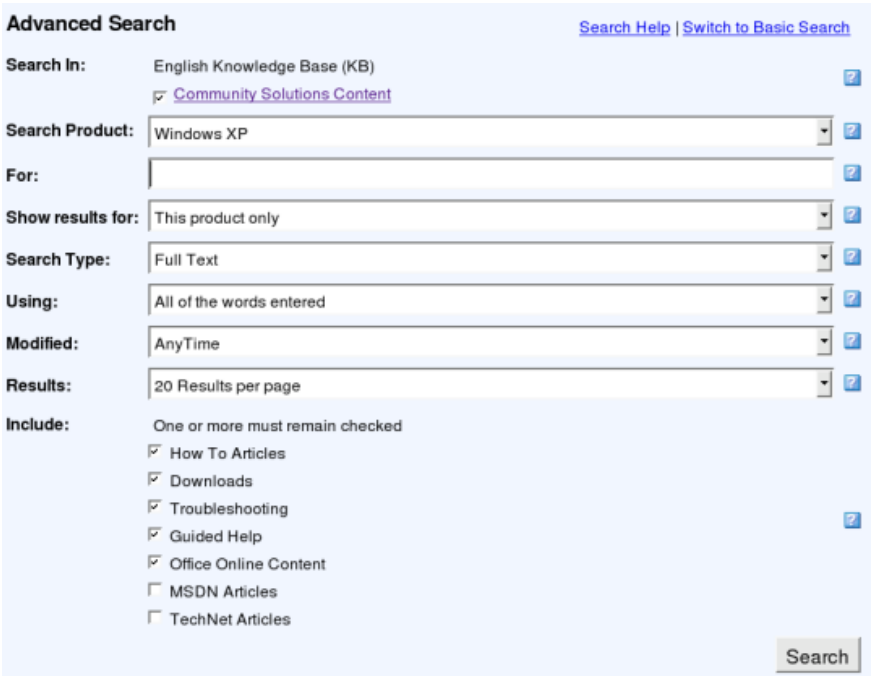

Figure 1: Microsoft support web site – Advanced search form

One of the reasons why no real "semantic" platform exist yet in the field may stem from the fact that the semantic approach and technologies remain still relatively new, compared to the tagging and "folksonomies" approach – folksonomies are taxonomies created collaboratively by a community of users. The cost of developing domain specific ontologies remain high, as presented in the section below focusing on collaborative ontology design.

In the area of commercial help desk products, some companies declare using ontologies for structuring the help desk knowledge base. The KNova product datasheet [27] states for instance that the product uses "an industry-specific ontology for technical support, operating system, networking and desktop application issues". The penetration of semantic

technologies into help desks remain however globally low: none of the other referenced products features ontology use for the support knowledge base management.

2.3.2. Existing mechanisms for converting questions and answers into reusable information

A key aspect of help desk systems relies on the mechanisms they provide for converting questions and answers into reusable information. This is particularly important in the case of domains characterized by very rapid changes, such as the Linux domain where new or updated software and hardware are made available several times a year, raising new issues and solving some of the existing ones.

The manual maintenance of "Frequently asked questions" is an old-time mechanism consisting of converting conversations into reusable information. As stated in the book "Smart mobs – The Next Social Revolution" by Howard Rheingold [14] "Informal social aggregation of useful knowledge goes back to the lists of frequently asked questions (FAQs) posted to some Usenet newsgroups since the 1980s. [...] Beyond their defensive function, FAQs constitute a new kind of encyclopaedia in themselves, collectively gathered, verified, articulated webs of knowledge about hundreds of topics."

With the advent of the Semantic Web, new tools emerge for helping community members in maintaining a well structured knowledge base building upon the conversations that take place. It is no doubt that the act of conversation is one of the key process for knowledge sharing and knowledge creation, as emphasized by D.Grey (KM Consultant, maintainer of KMWiki[2]): "Conversations enable experience sharing via story, metaphor and analogy, - key steps for gaining understanding, learning, sense-making and becoming aware. Conversations allow for social connection, permit the emergence of new connections and meaning, provide the medium for promise and commitment, set the stage for future actions and encourage questions that spark deeper reflections and reviews." The challenge for semantic systems lies in creating automatically an ontological representation of existing conversations that capture the core facts stated. It is important however to note that the ontological representation will never replace the conversation flow, which contains not only objective facts but a meaningful context.

Nonaka and Takeuchi's [1] knowledge conversion model distinguishes four modes of knowledge conversion: socialization, externalization, combination and internalization. These modes are represented in the figure below.

[2] http://kmwiki.wikispaces.com/

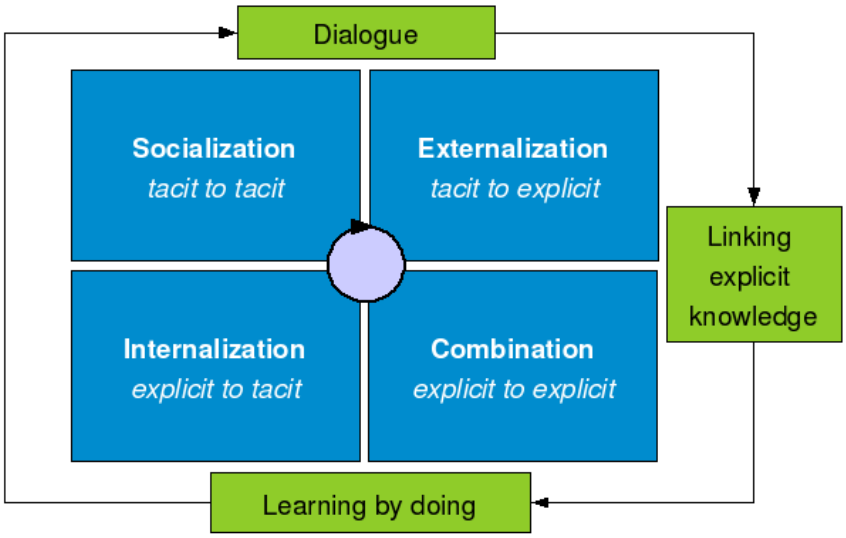

Figure 2: Nonaka and Takeuchi "SECI knowledge conversion model"

In Table 4, each mode of the SECI model is further explained in the context of the processes at stake within the current Mandriva Club community.

Knowledge creation mode	Description	Mandriva Club context
Socialization	Exchange of tacit knowledge among members that create common mental models and abilities. Socialization transfers tacit knowledge most frequently through the medium of shared experience. Apprentices learn by observation and imitation of the expert's behavior, as well as children, in ancient societies and, less frequently, nowadays.	Socialization takes place during "Linux install parties" where newcomers can bring their computer and discuss with experts volunteering for helping them in installing Linux and other open-source software
Externalization	The process of articulating tacit knowledge and transforming it into models, concepts, analogies, stories and metaphors that can be communicated by language. Externalization is considered to be a key phase in the creation of new knowledge and is induced by dialog, collective reflection, writing. Writing about knowledge is a good example of the effort normally required by externalization projects, as all humans have a whole life long knowledge experience.	Externalization takes place on the Mandriva Club forum, where users can submit questions. The Club forum is organized by language and by categories within each language. As of 2006, the following languages are hosted: English, French, German, Brazilian Portuguese, Spanish, Norwegian, Italian, Dutch.
Combination	Combination is the process of combining or reconfiguring bodies of existing explicit knowledge in order to generate new explicit knowledge, by addition. It is the most common process in formal education. In organizations, it is obtained by the exchange of explicit knowledge among members, as in formal reunions.	This process is typically missing currently on the Club. Building on Nepomuk core services, this process will be instrumented in the span of the WP11000 case study.
Internalization	Internalization is the process of adding to explicit knowledge (principles, procedures, methodologies) tacit new knowledge (in the form of sensations, memories, images) through experimenting in	This process exists but is not yet supported by appropriate tooling. It is the case study objective to support further the internalization process by providing the community with

Knowledge creation mode	Description	Mandriva Club context
	various ways, as through real life experience, simulation of limit situations or simulation through the usage of software.	tools that let add personal explicit knowledge to public resources.

Table 4: Nonaka and Takeuchi knowledge conversion model in the context of the Mandriva Club community processes

The "Combination" stage is the one that directly relates to the conversion of online discussions through forums or emails to semi-structured reusable knowledge. Putting into practice the components developed within Nepomuk core workpackages, WP11000 will provide a mechanism for associating a semantic representation with each discussion, so that the help desk partakers will easily convert these discussions into pieces of knowledge that can be found and reused easily.

2.3.3. State of the art on collective ontological design

Every ontology is a treaty – a social agreement – among people with some common motive in sharing.

Tom Gruber, Intraspect Software, http://tomgruber.org/

We believe that [centralized ontology building] efforts are unsustainable and that the Semantic Web will eventually be built in the same way as the WWW was – by its users

Good and Wilkinson, "The Life Sciences Semantic Web is Full of Creeps!", Briefings in Bioinformatics

Projects in this domain are premised on the assumption that, by distributing the burden of knowledge representation over a large number of people simultaneously, the knowledge acquisition bottleneck can be avoided

Benjamin M. Good, Erin M. Tranfield, Poh C. Tan, Marlene Shehata, Gurpreet K. Singhera, John Gosselink, Elena B. Okon, and Mark D. Wilkinson - Pacific Symposium on Biocomputing 2006

WP11000 case study domain relates to the management of software and hardware issues experienced by users running a Mandriva Linux operating system. The contents, WP11000 is dealing with, will be described using several ontologies covering the areas of computer hardware and of open-source software. These are wide and complex domains for which the design of dedicated ontologies is likely to require the input of many experts. In order to avoid well identified pitfalls while drafting these targeted ontologies, it is necessary to assess the current state of the art in the field of collective ontology design.

Collective ontology design is a relatively new field on the web. In an article that has been much commented on within the blogosphere, Clay Shirky [2], a well known consultant and teacher at NYU specialized on the social and economic effects of Internet technologies, points out that the cost for creating and maintaining ontologies pertaining to evolving domains is so high that the interest of ontologies may be overrated.

Clay Shirky claims that "ontologies will work well for domains that have a small corpus, formal categories, stable entities, restricted entities, clear edges." The Mandriva Linux community realm does not meet these expectations, quite the opposite: the corpus of documents to be

described is comprised of several thousands of items and the categories for describing hardware and software evolve at rapid pace.

Along the same lines, David Shotton and Chris Catton from the University of Oxford [3] put forth the pitfalls of ontology design, stating that:

"Ontology building requires a high level of understanding, time and dedication, and most ontologies are actually build by a small dedicated group of 'monks'. True community involvement in ontology building is thus difficult, and is often more of a platonic ideal than a reality. The ontology-building 'monks' are usually led by an 'abbot', a relatively senior domain expert likely to be committed to encapsulating the dominant paradigm. Substantial logistic problems confront any newcomers wishing to contribute, since ontology building is time-consuming and thus expensive in manpower. Since an ontology expresses the community consensus, there will be massive social pressures against change".

Shotton also warns for the fact that "social and technical factors favour the fossilization of current domain paradigms into static ontologies".

This vision emphasizing the dreadful cost of ontology design is however counterbalanced by other arguments and by real world successful experiments that are mentioned below.

While directly responding to C.Shirky's statements, Nova Spivack, founder of Radar Networks[3], coins the term "folktology" and presents the "folktology" approach as the next pattern in collective knowledge modelling [4]:

"Imagine a folksonomy combined with an ontology - a 'folktology'. In a folktology, users could instantly propose or modify ontological classes and properties in the same manner that they do with tags in tagging systems. The most popular ontological constructs (the most-instantiated classes, or slots on classes, for example) would "rise to the top" and self-amplify, while the less-instantiated ones would "fall to the bottom" over time. In this way an emergent, self-organizing, and self-pruning ontology could emerge within a community. Such a system would have the ease and adaptability of a folksonomy plus the semantic richness and formal structure of an ontology. I think ultimately a folktology approach will be better than either folksonomies or ontologies on their own."

In a presentation given during a Microsoft Research Faculty Summit 2006, Benjamin Good from University of British Columbia, introduces successful folktology experiments [5]. While assessing the cost of large ontology design, Good first expresses the same concern as Shotton and Shirky, and even questions the affordability of the Semantic Web in that context:

"Gene Ontology: Curated: ~5 full-time staff, $25 Million (Lewis,S personal communication)

National Cancer Institute Metathesaurus: Curated: ~12 full-time staff, $75 Million (personal estimate)

Health Level 7 (HL7): Curated – staffing unknown, $15 Billion(?) (Smith, Barry, KBB Workshop, and Montreal, 2005)

---

[3] Radar Networks: http://www.radarnetworks.com/

To build the Semantic Web for Life Sciences we need to encode knowledge from EVERY domain of biology – from barley root apex structure and function, to HIV clinical-trials outcomes... and this knowledge is constantly changing! At >>$25M a pop, can we afford the Semantic Web???"

Good however draws the opposite conclusion of Shirky's, introducing the methodology dubbed "iCapturer", based on community involvement for producing accurate ontologies at relatively low cost. The iCapturer methodology is presented as "a paradigm shift in knowledge capture methodologies": iCapturer consists in an "open, parallel, decentralized, synergistic protocol" at the opposite of the traditional "centralized, highly curatorial model employed in the development of all of the major bio/medical ontologies produced to date". The methodology has been experimented with several user communities. One of the experiments is presented in the article "Fast cheap zero cost curation model for ontology building", the abstract of which stating that in "two days at a conference focused on circulatory and respiratory health, 68 volunteers untrained in knowledge engineering [...] created a shared vocabulary of 661 terms, linking these terms to each other and to a pre-existing upper ontology by adding 245 hyponym relationships and 340 synonym relationships [...] using a web-based interface." The article further describes the protocol used and presents "quantitative and qualitative assessments of the constructed ontology" [21].

Along the same lines, in the article "Harvesting Wiki Consensus - Using Wikipedia Entries as Ontology Elements" [22], Martin Hepp, Daniel Bachlechner, and Katharina Siorpaes present how the contributors to Wikipedia manage to reach a consensus in creating ontologies collaboratively:

"We can observe that, within Wiki communities, especially Wikipedia, a large number of users is able to create comprehensive domain representations in the sense of unique, machine-feasible, identifiers and concept definitions which are sufficient for humans to grasp the intension of the concepts. [...] We (1) show that standard Wiki technology can be easily used as an ontology development environment for named classes, reducing entry barriers for the participation of users in the creation and maintenance of lightweight ontologies, (2) prove that the URIs of Wikipedia entries are surprisingly reliable identifiers for ontology concepts, and (3) demonstrate the applicability of our approach in a use case."

Beside wikis, several tools are available for collaborative ontology design. A study by Yiling Lu presents a comparison of existing tools, in "Roadmap for Tool Support for Collaborative Ontology Engineering" [23]. The study focuses on five ontology authoring tools with a specific focus on collaboration support: Ontolingua server, OntoEdit, Apecks, CO4 system and Protégé. In the end, while recommending Protégé and JXTA for collaborative authoring of ontologies, the study concludes that "collaborative ontology development is not well supported by any of the existing ontology authoring tools or environments." This somewhat drab conclusion has to be tempered however by the fact the study was issued three years ago, in 2003.

Another encouraging argument for the potential success of the "folktology" approach is put forth by Tom Gruber, who stresses the fact that ontology engineering is in the end very close to software engineering [24]:

"It seems to me that ontological engineering faces the same problems as software engineering. I would look for ontological engineering methods

that address the issues of functional specification (what the ontology is for), design documentation and review, enforcing constraints while editing, unit and integration testing, and the distributed collaboration that would allow the ontologies to be developed in an open-source style."

From this statement and since Mandriva Linux community has a long experience in collaborative software engineering, the proximity of software and ontology engineering let us expect a successful outcome in the field of ontology engineering within WP11000.

There remain however open problems in the field of ontology engineering: Shotton [3] points out that "it is difficult to record the provenance of ontological change. It is impossible to encode and use different (old and new) versions within the same ontology, enabling the old version to be employed to interpret legacy data recorded using it, thereby avoiding a proliferation of separate ontologies." Shotton proposes the use of "Named Graphs to permit provenance and other metadata to be attached to subgraphs within an ontology", and the development of "LiveOWL, an extended version of OWL incorporating Named Graphs". The potential need for such an extension within WP11000 and the current status of LiveOWL will be further assessed during the lifespan of WP11000.

2.4. Social axis

Beside the innovation brought by the semantic approach, Nepomuk platform in general and WP11000 in particular aim at harnessing new knowledge sharing paradigms through social networking.

The social aspects of the online community platform targeted by WP11000 relate to (i) methods to rate the quality of a contribution (ii) methods to establish a trust network across the community members (iii) methods to raise the quality of user contributions (iv) methods for harnessing collectively a common knowledge base. The paragraphs below describe existing methods or research efforts aiming at addressing these issues.

2.4.1. Methods to rate the quality of a contribution

Word-of-mouth, one of the most ancient mechanisms in the history of human society, is being given new significance by this unique property of the Internet. Online feedback mechanisms, also known as reputation systems (Resnick et al., 2000), are using the Internet's bi-directional communication capabilities in order to artificially engineer large-scale word-of- mouth networks in which individuals share opinions and experiences on a wide range of topics.

Chrysanthos Dellarocas, The digitization of word-of-mouth: promise and challenges of online feedback mechanisms [7]

In order to assess collectively the value of contributions, many online community platforms are equipped with mechanisms that let users provide feedback to the contents submitted by others. Table 5 describes the solicited feedback on two platforms, Google Answers and Yahoo! Answers that leverage most of this type of mechanism.

Platform	Format of solicited feedback	Remarks
Google Answers	Customers can rate answers from one to five stars. Customers can also refuse the payment if they're not satisfied with the answer	Researchers who receive too many negative reviews may have their privileges revoked by Google.
Yahoo! Answers	Community members can provide positive or negative feedback on any question or answers as long as they've reached the "level 2" or up contributor level (details explained below). All members can also report abuse. Askers can rate positively or negatively a question.	

Table 5: Format of solicited feedbacks on existing help desk platforms

2.4.2. Contributor rating and reputation management

Reputation marks the spot where technology and cooperation converge.

Howard Rheingold, Smart Mobs – The Next Social Revolution [14]

The web is more a social creation than a technical one. I designed it for a social effect — to help people work together — and not as a technical toy. The ultimate goal of the web is to support and improve our weblike existence in the world. We clump into families, associations, and companies. We develop trust across the miles and distrust around the corner.

Tim Berners Lee, Weaving the web [15]

Contributor rating generally relies on two indicators: one relates to the number of contributions while another relates to the qualitative score received from others. The latter can be either the average of several scores received for each contribution, or a mean of scores on a given area of expertise received from several members based on the view they have of the user.

In most platforms, users end up being rated through their contributions: when user A rates an object created by user B, user B's rating will automatically be impacted. Three levels of global qualitative rating can be identified, independent of the transaction:

- Level 1: the platform will "tag" the user (Featured contributor of the month, Top expert, #2 in Top 10, etc.)
- Level 2: the user will tag himself (PHP expert, Linux newbie, etc.)
- Level 3: the user is tagged by others (Genius, efficient, impolite, etc.). Instead of writing a full-text comment, a set of tags can be enough to convey the same content while making it easier for the users and the system to treat and visualize.

2.4.3. Social methods to sustain and raise the quality of user contributions

In the article "Motivating participation by displaying the value of contribution", Paul Resnick et al. explain how the public availability of each members' rating and contribution history is a key aspect in encouraging user contributions [25].

Several systems have been adopted by existing help desks for conveying special recognition to the best contributors. Recognition either relies on "ego boosting" through point systems and a hall of fame of the most active and best rated contributors, or on commercial partnerships.

Qunu platform displays in a convenient way the statistics related to a user's contribution, as shown in the picture below. Two key indicators are displayed: user rating by domain (Wordpress software domain in the example below), and contribution history (number of contributions in this case).

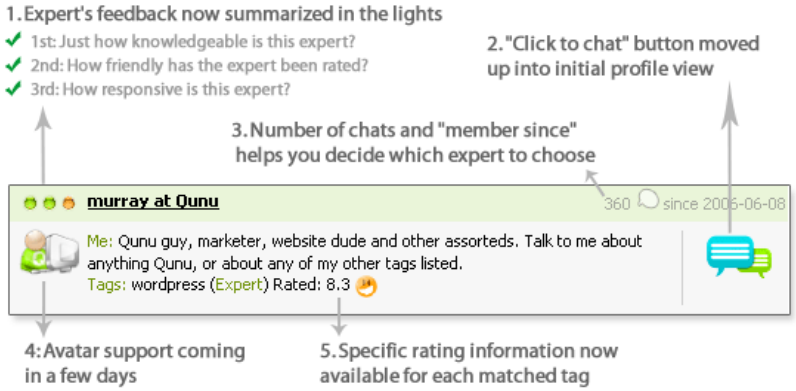

Figure 3: Expert information summary on Qunu

Yahoo! Answers is also putting forth the rating of each user, providing following incentive for contribution:

"To encourage participation and reward great answers, Yahoo! Answers has a system of points and levels. The number of points you get depends on the specific action you take." Yahoo! Answers insist on the social value of the awarded points: "While you can't use points to buy or redeem anything, they do allow everyone to recognize how active and helpful you've been. (And they give you another excuse to brag to your friends.)"[4]

Yahoo! Answers scoring system is further described in the table 6.

Microsoft QnA scoring system follows a path very similar to Yahoo! Answers, relying also on ego boosting: "We want to give QnA members who give to the community the respect (and the glory) they deserve. There are several ways to get famous and trusted on QnA."[5]

As far as Microsoft professional products is concerned, Microsoft has set up a more formal rewarding process in creating "The Microsoft Most Valuable Professional (MVP) Program". Microsoft MVPs are volunteers who have been awarded for providing technical expertise towards communities supporting Microsoft products or technologies. Microsoft describes the program as follows:

[4] http://answers.yahoo.com/info/scoring_system.php

[5] http://qna.live.com/CommunityContent.aspx?frame=scoring

Action	Points
Begin participating on Yahoo! Answers	One time: 100
Ask a question	-5
Choose a best answer for your question	3
No Best Answer was selected by voters on your question	Points Returned: 5
Answer a question	2
Deleting an answer	-2
Log in to Yahoo! Answers	Once daily: 1
Vote for a best answer	1
Vote for No best answer	0
Have your answer selected as the best answer	10
Receive a "thumbs-up" rating on a best answer that you wrote (up to 50 thumbs-up are counted)	1 per "thumbs-up"

Table 6: Yahoo! Answers scoring system

"The Most Valuable Professional Program is the way that Microsoft formally acknowledges the accomplishments of these individuals for their contributions to community. [...] The key strategies the program employs are:

- Recognize and engage with MVPs worldwide—Identify, enable and empower community influencers through a consistent quality customer relationship with Microsoft that spans product groups, services, and field organizations.
- Improve customer connection and satisfaction—Recognize more customers for their efforts and improve the quality of the experience on their turf and in their language.
- Drive program excellence—Professionalize services, customer offerings and worldwide roles and responsibilities to become more predictable and accountable to both internal and external Microsoft community customers.

The MVP Program, in existence for over eleven years, is represented by over 2,600 MVPs in 81 countries."[6]

Several web sites are devoted to the MVP Community. Within a dedicated section of the main Microsoft Support web site[7], the "MVP of the month" is interviewed in a column called "MVP Insider"[8]. Microsoft also organizes on-line press coverage of MVPs activities[9].

The detailed profile of all current MVP awardees is available on-line at http://mvp.support.microsoft.com/communities/mvp.aspx, where MVPs can be browsed by their area of expertise (more than 100 specific

[6] http://mvp.support.microsoft.com/mvpexecsum

[7] MVP web site: http://mvp.support.microsoft.com/

[8] MVPINSIDER: http://mvp.support.microsoft.com/MVPINSIDER

[9] MVP press coverage: http://mvp.support.microsoft.com/mvpprsc

technical areas are listed on the page). The MVP profile consists of a photo, a biography abstract, a list of publications, speaking engagements, conferences, professional position, a list of recommendations, language competencies and contact information.

This vibrant community of contributors is also sustained by external sites such as http://msmvps.com/ "The Ultimate Destination for Blogs by Current and Former Microsoft Most Valuable Professionals" and http://www.mvps.org/ "Your jumping off point to a number of interesting offerings being provided for you by a few folks associated with the Microsoft Most Valuable Professional program".

It is important to note that, as stated in the FAQ section of the main MVP web site "MVPs are not employees of Microsoft nor do they speak on Microsoft's behalf. MVPs are third party individuals who have simply received an award from Microsoft", and that "MVPs do not receive any monetary payment from the Microsoft MVP Program."[10]

2.4.4. Methods for social knowledge work

In his book "Smart mobs – The Next Social Revolution" [14], Howard Rheingold emphasizes the power of platforms letting users form groups, referring to the Reed law:

"Reed's law is the link between computer networks and social networks. Reed, using his law to analyse the value of different kinds of networks, believes he has discovered an important cultural and economic shift. When a network is aimed at broadcasting something of value to individuals, like a television network, the value of services is linear. When the network enables transactions between individual nodes, the value is squared. When the same network includes ways for the individuals to form groups, the value is exponential."

The value of different types of networks is, by Rheingold, summarized as follows:

Social network type	Network value
Network dominated by linear connectivity value growth	Content is king.
Networks where Metcalfe's law dominate	Transactions traded stuff is king.
Group forming networks	Central role filled by jointly constructed value (specialized newsgroups, joint responses to RFPs etc.).

Table 7: Comparison of different types of network values

Along these lines, Yahoo! features a "My Web" service that lets individuals save bookmarks and notes into a personal space that can be shared with others. This space is directly accessible from Yahoo! Answers. Groups can hence easily be formed on the Yahoo! platform. Shared spaces can then be searched by participating members: this is called "social search".

[10] MVP FAQ: http://mvp.support.microsoft.com/default.aspx?scid=fh;EN-US;mvpfaqs

Social search is a new type of search taking advantage of social networks. Even if the concept is not yet directly exposed in existing online help desk systems such as Yahoo! Answers, the combination of community search and community help desk is likely to be of great interest to users while they search for help.

Yahoo! social search features are further described in a CNet article titled "Yahoo! tests 'social' search"[11]:

"Dubbed My Web 2.0, the service builds on personalized search features [...] allowing Yahoo users to archive their search results and share them with other people using the service, but the next iteration will go even further. People with a Yahoo login will be able to bookmark and cache copies of their favorite Web sites, label them in certain categories and attach comments in a structured way. Users will then be able to search among their contacts' knowledge base with what Yahoo is calling its MyRank search technology. [...] But typically do a poor job of connecting you with new items that might be interesting, timely and personally relevant. Your friends and people who share common interests with you are better sources for this information. [...] Over time, we envision communities using My Web to build their own search engines to capture and make accessible the knowledge of their community"

Swicki[12] and Wink[13] are other examples of social search engines that have been gaining interest since their launch in 2006. Swicki engine is described as "A search engine that learns from your community's search behavior. Automatically."

In a project proposal that has been awarded by Microsoft Live Labs, Gerd Stumme predicts a bright future for such social search tools: "Unlike link-based search approaches à la PageRank, these systems provide personal recommendations based on input from similar users. This new paradigm will change the way we are interacting with the web within the next few years. In particular, it will require corresponding search functionality. Furthermore, these systems are more responsive to upcoming topics, which can thus earlier be discovered and actively promoted. Therefore, we will extend link-based search with social search, in order to provide enhanced functionality and multiple search paradigms for the Web." [26]

2.5. Desktop integration axis

Since the workflows at stake in help desking are potentially complex and time consuming, an in depth integration of the help desk features into the desktop is another important aspect for increasing the productivity of Mandriva Linux knowledge workers.

As of 2006, online platforms such as Google Answers, Yahoo! Answers and Microsoft QnA remain purely web applications with no desktop

[11] Yahoo! tests 'social' search: http://news.com.com/Yahoo+tests+social+search/2100-1038_3-5768135.html

[12] http://swicki.eurekster.com/

[13] http://www.wink.com

extensions. Open-source help desk systems such as OTRS are also purely web based. Kayako Support Suite however features an integration with Microsoft Outlook and provides several desktop add-ons such as the Kayako InstaAlert: "Kayako InstaAlert allows you to receive real-time alerts whenever a ticket gets updated under the assigned departments. The application displays popups as and when the tickets are created or replied to allowing you to answer your customer requests and issues promptly."[14]

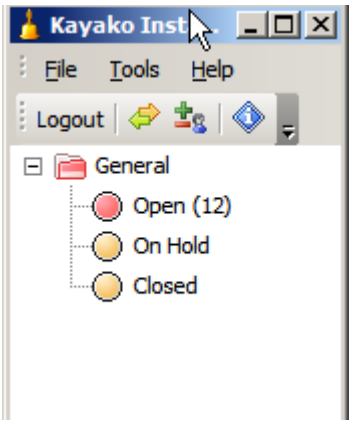

Figure 4: Kayako InstaAlert

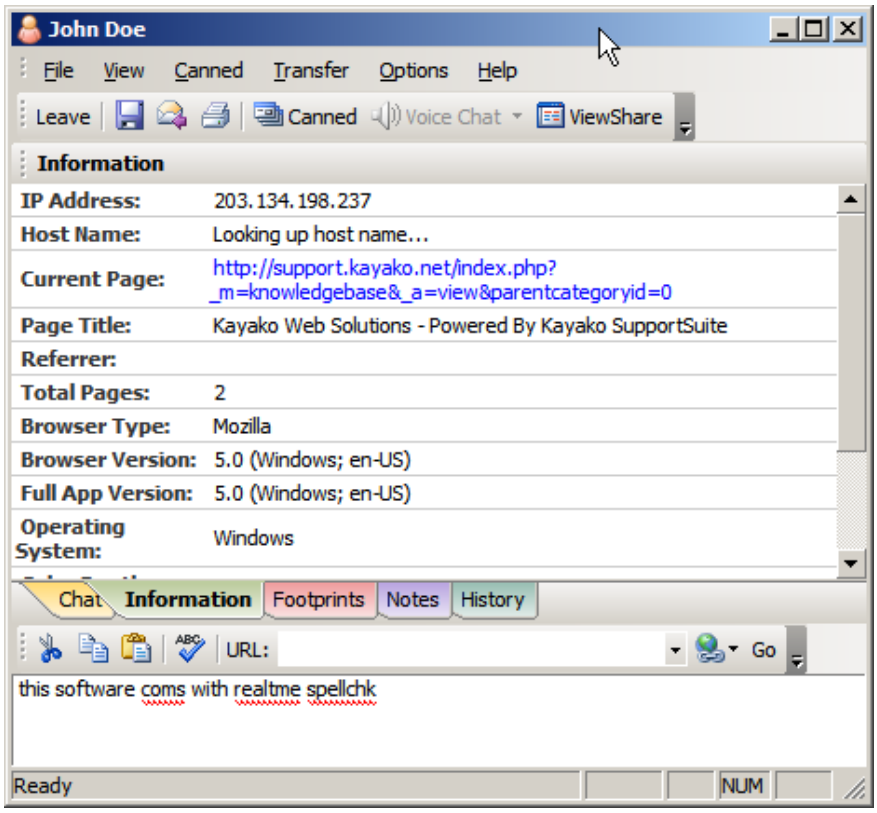

Figure 5: Kayako desktop chat window, including a live spell check

[14] http://www.kayako.com/instaalert.php

Qunu help desk system uses Jabber protocol for communication, hence letting members use their favourite Jabber desktop client for chatting.

Qunu platform also recommends the use of a number of satellite applications for facilitating the exchange of data between the experts and the users. These applications are meant to enhance the support that is possible via Qunu. All these tools remain however web applications with no real integration with the desktop.

Qunu Support Tool	Description
Pastebin (http://pastebin.ca)	Useful for passing large blocks of text, code, logs, etc. Provides formatting and syntax highlighting for many popular file types.
Encoding tools http://opinionatedgeek.com/	Several tools for encoding/decoding Base64, HTML and UU formats
PrintKey Screenshot program http://www.321download.com/LastFreeware/page4.html#PrintKey	A useful tool to take screenshots of everything; doesn't need install, is just a standalone program
Screenshots hosting http://photobucket.com http://flickr.com http://imageshack.us	Good for sharing screenshots
Screen movies http://www.camstudio.org	Utility to record and encode desktop activity
file transfers http://www.dropload.com/	Great for sending big files
Remote desktop http://www.ultravnc.com/	Customizable utility to easily access a remote desktop through VNC without installation
Remote desktop	Example: mstsc

Table 8: Qunu support tools for enhanced data exchange

The tools above suggested by Qunu web site are good candidates to be integrated into the WP11000 desktop application, so that data exchange between users and experts gets as simple as possible, with no need to use several distributed applications over the web for exchanging data.

2.6. Openness and extensibility

Beside the core functional features of the help desk platform, the advent of Web 2.0 mashup applications is emphasizing the fact that open and extensible frameworks considerably favours community uptake.

Yahoo! Answers for instance provides the developers with a set of Web Services APIs harnessing the power of the platform[15]. As of 2006, Yahoo! Answers only provides REST web services, not SOAP web services yet.

Table 9 presents the services currently accessible from the Yahoo! Answers platform.

[15] Yahoo! Answers API home page: http://developer.yahoo.com/answers/

Web Service Name	Web Service Description
questionSearch	Find questions that match your query
getByCategory	List questions from one of hundreds of categories, filtered by type. You'll need the category name or ID, which you can get from questionSearch.
getQuestion	Found an interesting question? getQuestion lists all the details for every answer to the question ID you specify, including the best answer, if it's been chosen. Get that question ID from questionSearch or getByCategory.
getByUser	List questions from specific users on Yahoo! Answers. You'll need the user id, which you can get from any of the other services listed above.

Table 9: List of Yahoo! Answers Web Services available through REST calls

All services can be tested live from Yahoo! web site through a form. The services are however limited to 5,000 queries per IP per day per API.

Each service is very well documented on Yahoo! web site. The documentation includes the list of all available parameters and the data schema of the response sent. While using the “questionSearch” service for instance, developers can specify the question category, its date, the text to be searched for (either in questions, answers, nicknames etc.), the desired response output type (which can be either XML, JSON, PHP or RSS) etc.

Yahoo! has also opened a "Yahoo! Answers application Gallery"[16] featuring the best applications developed on top of the provided API.

To the best of our knowledge, neither Google Answers nor Microsoft QnA offer a comparable web API yet.

2.7. Conclusion

The state of the art has shown that even though there is no real "social semantic help desk" implementation yet, the idea of combining Semantic Web technologies with social mechanisms for creating high quality data, meta-data and processes usable by large online communities in their effort to solve problems collectively is a work in progress within several of them.

A few communities have been experimenting semantic wikis for creating ontologies collectively. Given the power of the approach, it is likely that it will blossom at rapid pace both in content oriented communities and in software oriented communities (the existence of a Workshop on Semantic Web Enabled Software Engineering at ISWC[17] is a significant initiative in that direction).

[16] Yahoo! Answers application gallery http://gallery.yahoo.com/answers/

[17] http://www.mel.nist.gov/msid/conferences/SWESE/

Nepomuk WP11000 challenge will consist in delivering an appealing social semantic help desk in 2007 building upon Nepomuk core services. A key factor of success will rely on the Mandriva Linux community adoption of the system, both as help desk users and content producers. The in depth integration of the system with desktop technologies and rich user interfaces available from different technological framework will act as a key factor of adoption of the system among end users.

The effects of such a system on the management of knowledge and processes within the Mandriva Linux community are potentially huge. In easing dramatically the way end users can get precise answers to their questions, the help desk is likely not only to support the production and innovation processes of the Mandriva Linux community, but also to have a tremendous impact on the adoption of Linux based systems on the desktop.

3. Business case for knowledge management improvement within the Mandriva Community

This section describes the global environment of the Mandriva Club as a commercial service and the challenges the service is currently facing. The organizational approach for Nepomuk based help desk adoption within the Mandriva community is then presented, followed by a cost/benefit analysis.

3.1. Mandriva Club infrastructure history

The Club platform has been running since 2001. Since its inception, several technical evolutions have occurred: in 2001, the early system was entirely based on PhpNuke system with home-made enhancements. In 2003, a wiki-based knowledge base was introduced, based on PhpWiki, and the forum system became a web site of its own, building upon PhpBB forum engine. In 2005, the main platform was migrated to XWiki engine, an advanced wiki system that includes Enterprise Content Management features such as an access right service, a translation service, a service for structured objects handling. The goal of this move was to involve as much as possible the Club members in the production of new contents and in the global Mandriva Linux innovation process. This move has been successful, but the system can be improved in several respects. The objective of WP11000 is to put Nepomuk components into practice within the Mandriva Club platform for improving the knowledge sharing and production processes at stake in the Club.

Figure 6: *Mandriva Club home page (http://club.mandriva.com)*

The Mandriva Club consists of following main modules: a knowledge base, a forum (available in 6 languages), an e-learning module, a P2P download module, a blog module, a chat, a module for requesting new RPMs, and a calendar of Linux related events all over the planet.

The Club user interface is available in 25 languages.

3.2. Case study big picture

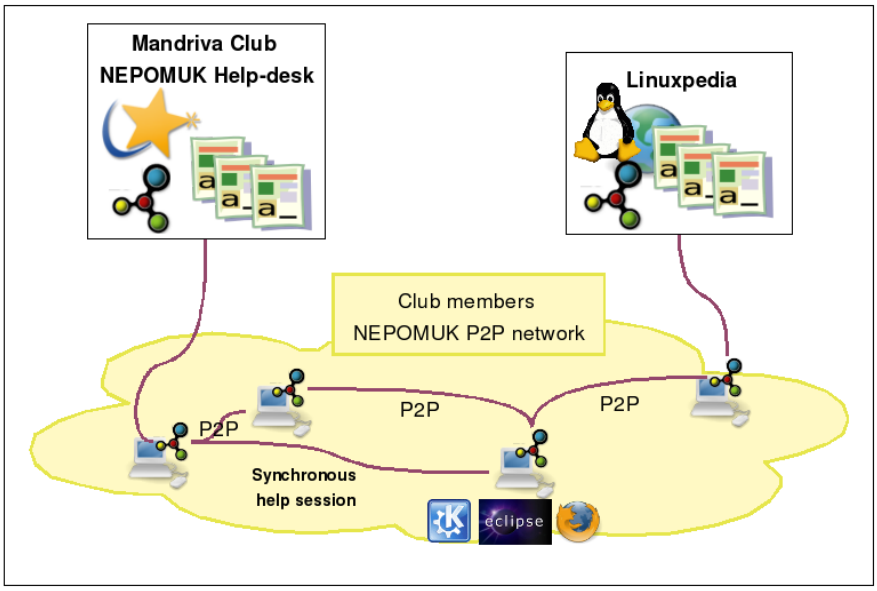

Figure 7: Mandriva Linux Community case study big picture

In the picture above, Mandriva Club members have a Nepomuk enabled desktop. This desktop includes rich user interfaces for manipulating information and for communicating over a P2P network. These components are based on the KDE, Eclipse and Mozilla Firefox frameworks and result from the involvement into the implementation of Nepomuk APIs by the corresponding developer communities, as devised by WP7000 dissemination workpackage. These users can share information across a P2P network connected to two main knowlege bases: Linuxpedia and Mandriva Club knowledge bases. The P2P network serves three purposes: first, the P2P communication enhances the information exchange process during live help sessions: users can share files, screenshots, command outputs and even voice messages directly from one desktop to another in live mode; second, it lets collaborative groups work on a set of documents that are shared on a private network of desktops without the need to transfer their private documents to a centralized area in the first place (this is especially useful when the shared documents amounts to a large size, or when a high security level is required); third it is used for downloading software packages through a BitTorrent like protocol. Users contributing with large bandwidth to the download system gain an increased karma.

The need for some kind of "Linuxpedia", i.e an encyclopaedia based on the organisational model of Wikipedia and focusing on learning material related to Linux and open-source software (OSS), has emerged from the user requirements analysis: it has appeared indeed that for maximizing community uptake in the production of semantic material on Linux and open-source software in general, an appropriate strategy may consist in bringing all open-source software actors together for producing and maintaining a distribution-agnostic knowledge base. In this context, the

Mandriva Club knowledge base would extend that common public OSS knowledge base, focusing on Mandriva Linux specific issues.

It happens the EU is funding the SELF project[18], which started September 2006. As outlined on the web site of the project, "SELF brings together universities, training centres, free software communities, software companies, publishing houses and government bodies to facilitate mutual support and exchange of educational and training materials on free software and open standards". The SELF platform may be an appropriate platform for the "Linuxpedia" envisaged by WP11000. Discussions with the SELF team will be conducted for assessing potential synergies between Nepomuk WP11000 and SELF.

In the scope of the case study, users can store on their desktop their own private annotations on resources stemming from Linuxpedia or from the Club knowledge base. These annotations constitute a key stage of the learning process, as emphasized above by the SECI knowledge conversion model by Nonaka and Takeuchi [1]. Users can also share their annotations and bookmarks with their acquaintances using the Nepomuk social layer.

The help desk system takes place on the Mandriva Club network and has strong links with the Mandriva Club and Linuxpedia knowledge bases. It lets users issue questions and bring answers to one another either in an asynchronous manner, or synchronously. This is illustrated by the two members at the bottom of the above illustration, who have initiated a live communication exchange so that one can bring help to the other on a specific topic that he has good expertise of, in a similar manner as on the Qunu platform.

3.3. Organizational framework for Nepomuk adoption within the Mandriva Club community

Since the involvement of the community is a key factor of success of the case study, it is crucial that the WP11000 team devotes time and energy for spurring the end users to take part in the design of the target system and to use it as their primary source for software/hardware problem resolution.

3.3.1. Promotion

The promotion of the system will consist in continuous interaction with as many members as possible. A dedicated mailing-list has been created at this prospect[19], and the Nepomuk team has created a blog on the Mandriva Club for keeping the community abreast of the system development and use[20].

---

[18] SELF Project: http://www.selfproject.eu

[19] Mailing-list: Nepomuk-mandrivaclub@Nepomuk.semanticdesktop.org

[20] http://club.mandriva.com/xwiki/bin/view/NEPOMUK/Blog

WP11000 team also plan to organize other workshops with Mandriva Linux users for evangelizing the Nepomuk approach in general and the use of the social semantic help desk in particular.

3.3.2. User training

Since some advanced features of the final system may require training, the WP11000 team will make sure the social semantic help desk comes with a full-fledged documentation manual. The WP11000 team will also organize training sessions.

3.3.3. Software evolution

The WP11000 team will set up a dedicated area within the Club for users to report feedback on the system, bugs and requests for enhancement. Edge-IT will carefully take into account these reports for continuously improving the system, in close cooperation with the core technical workpackages.

3.4. Costs benefits analysis

The design of those online platforms provides the frame in which knowledge is concentrated and activated as a resource for creation.

Andrea Hemetsberger and Christian Reinhardt
Sharing and Creating Knowledge in Open-Source Communities –
The case of KDE [6]

3.4.1. Expected benefits

As stated on Wikipedia, "Help desks play a key role in modern business organizations. A well designed and maintained help desk can substantially: increase the efficiency of operations, reduce costs, gain customer satisfaction, improve public image"[21].

Edge-IT expects many beneficial effects from the help desk. It is expected that the help desk will ease the maintenance of the knowledge base covering the use of software and hardware with Mandriva Linux. This social semantic knowledge base will be a key asset for sustaining the creativity of the Mandriva community as a whole: as stated in the article "Sharing and Creating Knowledge in Open-Source Communities – The case of KDE" [6], "The design of those online platforms provides the frame in which knowledge is concentrated and activated as a resource for creation." An efficient collaborative knowledge base where users easily will find an answer to their questions or a way to collaborate to existing content is likely to have direct impact on the adoption, the use

[21] http://en.wikipedia.org/wiki/Help_desk

and the spread of Mandriva Linux, as well as of open-source software in general. Even if the Linux desktop is getting more and more easy to use, the availability of a central place with up-to-date multi-lingual information organized in a way that takes advantage of Semantic Web technologies and that is maintained by a rich social network with member from all over the world will dramatically lower the entry barrier and the learning curve for open-source software (OSS) adopters in general, and for Mandriva Linux newcomers in particular.

The core services provided by the help desk will most probably be open to all members of the Mandriva Linux community. Edge-IT may however restrict some advanced services, such as advanced personalized search, to paying members of the Club.

The questions and answers and the associated knowledge base maintained by the community will also ease the production of up-to-date manuals addressing specific problems related to recent technologies or hardware. A dedicated business model could follow the "Extreme Publishing" approach described by Leander Kahney and Adam Engst (creator of Tidbits.com) in a Wired News article titled "Net Publishing Made Profitable"[22]:

"Edited collaboratively over the Net, the books are published 'within moments of going to press' as small, downloadable PDF files. Costing $5 or $10, the books come with free updates for readers - the electronic equivalent of second and third editions."

It is expected that active contributors to the help desk and to the knowledge base will form expert groups for producing integrated how-to manuals encompassing a well defined domain such as "USB devices management with Mandriva Linux" or "WIFI management with Mandriva Linux" etc. A possibility would consist in letting Mandriva Club members access freely these manuals and in offering them to the general audience for a fair price. The most active contributors would be financially compensated proportionally with the revenues generated by their manual.

A third line of revenue based on the final system will consist in creating partnerships with service companies for adapting the help desk solution to other companies with similar needs in different contexts. Provided the importance help desk systems are gaining on the web, a "social semantic help desk" software appliance may meet needs on the market, requiring specific adaptations to the idiosyncrasies of the targeted domain.

Last, an ecosystem of commercial actors can emerge from the community of partakers: in a way similar to what Google Answers proposes, some questions may be paying questions, for which individual experts or commercial companies would bring answers.

[22] http://www.wired.com/news/mac/0,2125,64563,00.html

3.4.2. Expected costs

Edge-IT expects the "crowdsourcing" approach to make it possible to have low operating costs and low investment capital as far as the maintenance of the content is concerned.

Low operational costs rely on the following premises:

- quality help is provided as much as possible by the users themselves.
- users are rewarded for their contribution, through means other than direct financial compensation when possible.
- the help desk sustains growth in the number of users demanding help, while maintaining and raising wherever possible the overall quality of provided help
- the help desk works with different languages and cultures

The development of the platform will be partially supported by the Nepomuk project. In the long run, the system will be maintained by several communities: on the one hand, it is expected that the SemanticDesktop.org community will keep maintaining the core components developed by Nepomuk; on the other hand, Edge-IT plans to coordinate a community of software developers working specifically on the integrated help desk system based on Nepomuk core components.

3.4.3. Balance

The shift to semantic indexing, amplified by the power of social networks is seen by some actors in the industry as a major shift in the way human and computers deal with information. Douglas Lenat, CEO of CyCorp states that "Once you have a truly massive amount of information integrated as knowledge, then the human-software system will be superhuman, in the same sense that mankind with writing is superhuman compared to mankind before writing."[23] On his side, Marc Fawzi, a venture capitalist blogging on Semantic Web and P2P technologies, sees the advent of combined semantic and P2P technologies as a major shift in the collective management of knowledge: "The availability of standardized ontologies that are being created by people, organizations, swarms, smart mobs, e-societies, etc, and the near-future availability of P2P Semantic Web Inference Engines that work with those ontologies means that we will be able to build an intelligent, decentralized, "P2P" version of Google"[24].

These statements tie in with the vision of a potentially considerable impact of a "social semantic help desk" on the adoption of Mandriva Linux, and even more important on the improvements of the Mandriva Linux system through collective problem solving and collective specification design of future versions.

[23] http://www.cyc.com/company

[24] http://evolvingtrends.wordpress.com/2006/07/11/p2p-search-the-peoples-google/

At the heart of Mandriva Linux project, Edge-IT is acting as a coordinator in information management, software specification design and software development. Even though it is difficult at this stage to evaluate precisely the cost/benefit balance, it is believed that the social semantic approach applied to the collective management of the information and processes related to the Mandriva Linux project can bring tremendous improvements outstripping the development costs.

4. User requirements

This section derives user requirements for Nepomuk WP11000 from the business case and from user observations. The section first presents the adopted approach for getting to know better the target users through interviews, surveys and workshops. A set of personas representing typical users is then introduced. These personas illustrate a series of scenarios putting forth what a social semantic help desk at work would mean. In accordance with these scenarios, the case study actors, data structures and processes are then described. The scenarios also lay the ground for defining the functional requirements of the help desk, as well as the non functional requirements. The last section presents the preliminary specifications of the domain ontologies that will be designed within the case study.

4.1. User observations

Members of the Mandriva Linux community are located all over the world and their activity (e.g., searching for information, contributing by answering questions in the forum, looking for new downloads, etc.) in the community varies. Mandriva users can either be members of different levels or non members of the community. Mandriva also has a few people employed with the role of contributing to the information on the Club (e.g answering questions on the forum).

To find out more about how the members use the community facilities provided by Mandriva, what kind of problems they encounter, and what their needs for a semantic help desk are, we have conducted interviews, distributed a questionnaire, organised a workshop, and developed scenarios and prototypes.

4.1.1. Interviews

Based on the memberships in Mandriva Club we have contacted people living in Sweden, Norway, UK, and Ireland for an interview. In total we have conducted ten interviews, among which six in Sweden, one in Norway, two in UK, and one on Ireland. Three of the interviews have been conducted in the respondent's home. The others have been conducted in an office or café. All interviews except one have been made in combination with the user demonstrating how he uses the Mandriva Club web pages. All interviews except one have been recorded, either using video camera or an iPod.

All members interviewed are male, between 30 and 65, and have a family (all but one had at least one child). They all have a passion for the open-source community and they use Linux as an alternative to Windows for financial or ideological reasons. They see their own contributions to the Mandriva Club, both the membership fee and information contributions, as a way of supporting the community. The respondents are not necessarily faithful to only using Mandriva Linux. Some use several distributions (e.g., Ubuntu, Suse, etc.), some use only Mandriva Linux, and some also use Windows in parallel for different reasons. Most of the respondents also have not only one or two computers, but several: one for running backups, one running as a server, one for the kids, one with another Linux distribution, and so forth.

Also, Mandriva Linux users appear to have a genuine interest in some kind of technological issue, in a homo faber way, e.g., building computers from different components, exploring different operating systems, working with radio amateur technology, trying out movie streaming, using sensor technology, or working with software development. Although most of the respondents has a genuine technical interest, their computer skills differ. Some of them have used Linux for a couple of months, while others have used it since the beginning of Linux.

The Mandriva Club is mainly used by the respondents for downloading new software. It is also used to search for information, e.g., about new downloads, or to solve a problem, but many of the respondents felt that the community is too small to give proper answers to all questions. Some of the respondents find it difficult to find information on the Club, that the search engine could be better, that the forum could be better structured, and/or that there should be information integrated from different sources on the Club. Several of the respondents have asked for a "how-to", not only for newbies but also for quite advanced instructions.

Typical problems that Mandriva users encounter may be Mandriva specific, Window manager specific, or kernel related. A Mandriva specific problem is related to the distribution, concerning, e.g. sound, network or component installations, or related to the products included in the distribution, e.g. the search engine Kat. A window manager specific problem may be related to, e.g. KDE or applications like KMail. Kernel related problems concern, e.g. drivers for specific devices.

Other information sources used when solving problems are LinuxQuestions.org, Wikipedia, Bugzilla, newsgroups, and other distributions' web sites, generic Linux forums, and/or Google. Most of the respondents also said that they were more active on the Mandriva Club when they were new members. Later on, they visit the Club mainly to get new downloads.

4.1.2. Questionnaire

To get more quantitative data about the Mandriva Linux users and to identify possible users to contact for an interview, a questionnaire was published on the Mandriva Club web page. The questionnaire was available not only to members of Mandriva Club, but to all visitors of the Club[25]. In total approximately 100 users filled out the questionnaire. Questions asked concerned statistical matters such as how long they have been members, how often they visit, search and contribute to the community, the different facilities offered though the community, and how they manage to solve their problems using the community. Other questions asked concerned why they use Linux, why they do or do not contribute and what would make them contribute, what kind of contributions they make, what parts of the community they like best and why, how the community can be improved, and suggestions to a new semantic help desk.

[25] Nepomuk questionnaire on the Mandriva Club: http://club.mandriva.com/xwiki/bin/view/Main/NepomukSurvey1

It is clear, from the questionnaire answers, that the respondents choose to contribute because they know about the problem, and that they want to help if it is simple and quick to do so. Little time, not knowing the answer, stupid questions, bad behaviour, or difficulties in using, for example, the knowledge base, may be reasons not to contribute with an answer.

On the question about how Mandriva Club could improve their community, people suggested that it should be open for everyone in order to increase the number of community visitors/contributors, an integration with other Linux websites, a possibility to post news, more interactivity, e.g., through a to-do help list, providing a hardware database with help on what hardware to buy, a guarantee to provide answers to problems, better visibility of new information like bug fixes, more people who have the role of contributors, moderating and updating the information, reducing the number of websites, and providing easy ways to propose ideas, distributions, etc.

4.1.3. Workshop

In early July we gathered Mandriva Club members living in Paris to participate in a workshop together with Mandriva Linux developers. In all, five Mandriva Linux users attended the workshop (two came late, and one left early), together with three developers/contributors from Mandriva. The participants were divided into two groups and each group developed one prototype. At the end of the workshop session the prototypes were presented and discussed in the whole group. The group discussions as well as the presentation of the prototypes were video recorded.

4.1.4. Prototypes

Prototypes in this early stage are used to provide the users with a tool to show, in action, what solutions they find relevant" [Schrage 1996]. "No abstract description that needs to be interpreted is necessary. The users develop prototypes based on their desires, problems, context, etc." In our case, two of the prototypes have been developed by users of the Mandriva community. The others have been developed based on the user studies (all four in the project). All prototypes, created using pen and paper, exemplify scenarios of how the user could interact with the system that is going to be developed.

Prototype 1: Creating a documentation page on the wiki from a forum discussion[26]

One of the prototypes developed during the workshop describes an activity where a forum discussion is moved to the documentation part in the community. The user can edit contributions by removing answers that are not relevant for the document. All authors in the discussion are shown to the user. A preview function makes it possible to look at the

[26] http://www.csc.kth.se/forskning/mdi/nepomuk/video/vp_create_documentation_page_on_the_wiki_from_forum_discussion.html

documentation before saving it. In the preview window the user can do more detailed editing of the text. Before saving the text, the user selects, from an interactive menu, under what categories the text should be visible, e.g., "hardware", "laptop", "Dell".

Prototype 2: Taking hardware profile into account when searching27

The other prototype developed during the workshop in Paris describes an activity where the hardware profile is taken into account when searching for information. During the search the system also suggests additional keywords to be used in the search. The search results include links both to Mandriva Club specific pages and to general Linux pages. The links are ordered by how often the link has previously been followed by a Mandriva user.

Apart from the two prototypes developed by users during the workshop, some additional prototypes have been developed that are relevant for the Mandriva community. From the list of prototypes28, the following are of interest for the Mandriva case.

Prototype	Description
Suggested-annotation import	The propotype provides an example of tagging information.
Tagging information	The prototype is of interest as an example of tagging documents found locally within Mandriva or external.
Ssearch-refining visualisation	The prototype is of interest as an alternative to visualise searches.
Ontology enhanced free-text search browse	The prototype was made based on the Mandriva case study results.
TMI-Athens-July	The prototype is of interest as an example of searching from several sources at the same time.
Taking hardware profile into account when searching	The prototype was made by Mandriva users and developers, showing how to take hardware profile into account when searching.
Documentation page creation from forum discussion	The prototype was made by Mandriva users and developers. It shows how to create a documentation page on the wiki from a forum discussion.
Search refined in several steps	The prototype is of interest as an example of creating and making a preview of free-text search queries.
GalwayDirkPaper	The prototype is of interest as an example of searching for information.
GalwayDirkPaperResults	The prototype is of interest as an example of creating one piece of text from several sources.
GalwayDirkPaperPresent	The prototype is of interest as an example of filering a search.

27 http://www.csc.kth.se/forskning/mdi/nepomuk/video/vp_taking_hardware_profile_into_account_when_searching.html

28 Prototype list: http://www.csc.kth.se/forskning/mdi/nepomuk/video.html

Email tagging	The prototype is of interest as an example of refining tagging though related tags.
Dirk_GettingIvolvedProject	The prototype is of interest as an example of using a timeline, in the Mandriva case this could be used for presenting the Mandriva project, what happens next, what has happened before, distributions, dependencies to drivers, etc.
OntologyPackage	The prototype is of interest as an example of searching through ontologies.
Nimbus-focus	The prototype is of interest as an example of focusing a search.

Table 10: List of prototypes description in the context of WP11000

4.1.5. Findings

It is clear, from both the interviews and the questionnaire, that many members are more active within the Mandriva community in the beginning of their Linux usage. The more experienced they are, the less they visit the community. The more experienced users typically visit the Mandriva community to download software. It is also clear that people use other Linux information sources available, either general or distribution specific, in order to find information. Many of the Mandriva Linux users also complained about the Mandriva community being small, i.e. with too few contributors. Hence, it is important to find means to make people, especially the skilled Linux users, more active in the Mandriva community. One example of such an attempt is through the use of "karma"[29], which evolved during the discussion in the Paris workshop. It is also important to notice that all interviewees are happy to contribute now and then to the community, i.e. the small community is not an effect of people not wanting to contribute.

It is also clear that many community members have problems finding the information they search for on the Mandriva community, either because the information is scattered in different places, the search facilities are not very advanced, or because there is too little information. A number of semantically focused solutions can be of help in these cases, e.g., using ontologies to support free-text search, providing pre-filled in information about the hardware used, including general Linux information in the search result, narrowing the search result by selecting specific types of information.

Among others, two ideas for making a better semantic Mandriva community evolved from the workshop discussion: karma and Linuxpedia. Karma has to do with strengthening the contributors' status within the community. The bigger karma a member has, the more he has contributed to the community. Linuxpedia, as described in the business

[29] Karma is a Sanskrit word "meaning action, effect, destiny. Karma is a sum of all that an individual has done, is currently doing and will do." (Source: http://en.wikipedia.org/wiki/Karma). The term was introduced into the world of online communities by the Slashdot platform (http://www.slashdot.org), which replaces the real-world word-of-mouth mechanisms with a reputation scores in order to provide incentives to all members to contribute positively to the community. The score of each individual is dubbed the "karma".

case section, has to do with the distribution-agnostic problems that Linux users often face. Linuxpedia focuses on collecting distribution-agnostic Linux information, through user contributions in the same manner as Wikipedia.

4.2. Personas

The relevant actors of Mandriva Club are introduced by personas. A persona is a fictitious person that represents a user group, in our case the users of Mandriva Club. The personas developed are based on the user studies made, and they are a detailed description and a visualisation of the users. They have a life, goals and scenarios where they fulfil their goals. They help focusing on the users during the design and give all stakeholders in the project a clear picture of the users' needs and requirements. The personas provide a condition for everyone in the project to have the same view of the users. They serve as a constant reminder of the users when used in the design work. The personas also "depersonalise" discussions on functionality and allow the designers to focus on designing for the personas [Cooper 1999, Cooper et al 2003].

For the Mandriva case we have developed two personas: Kim and André. Kim is a more recent Mandriva Linux user, who is not very skilled in searching for information, and who may not contribute with more complex information. André, on the other hand, is a skilled Mandriva Linux user, who develops open source software and knows all about searching for information.

Apart from the two personas there is also a specific role that is important to highlight: the paid contributor. Today, there are three persons who are paid to be active and contribute to the Mandriva Club by responding to questions in the forum and entering information in the knowledge base. However, it is important to find other means than money that can make more people contribute and be more active on the forum. Such means can be found in other communities, which should be further studied. André is a persona who, given the right means, would contribute more to the community.

Persona related scenarios consist of detailed descriptions of the personas' work and how they can conduct their work using the new system to be designed. The scenarios are inspirational and the goal is to describe the feeling of using the new system. For each persona, the corresponding user group and role is specified explicitly. The scenarios are described in the next section.

4.2.1. André

Figure 8: Portrait of André

Persona description

André lives in Karlstad, a small Swedish city close to the Norwegian border. He is 58 years old, married and has three grown-up children. He is a member of an aeroplane club and is certified for flying a small Cessna, which he likes do on holidays. He almost finished a PhD in mathematics long time ago, but had to drop it when he moved back to Karlstad for family reasons.

On workdays, André drives his old Opel to work. He's in a large defense industry company, in a group developing software for management of aircraft documentation. The system (which has both military and civilian customers) is SGML based, and runs on both Linux and Windows NT. For maintaining the Linux version, André uses a Mandriva Linux server.

In spite of the often secretive and closed culture of the defense industry, André has a passion for open source. Since he is benefiting from open source software almost on a daily basis, he is allowed by his employer to contribute actively in the community during work hours (as long as he is keeping company secrets in house).

André contributes to the Mandriva community when he can help others with advanced programming issues, and has even contributed some of the kernel code since he needed that for his own system to run smoothly.

He is hoping to get an early retirement offer, so that he can spend more time with his hobby, and perhaps take care of grandchildren.

Goals

- Current goal: Continuously improve the documentation management system
- Current goal: Have a smoothly and securely running Mandriva installation
- Long-term goal: Make his code maintainable and sustainable so that his successor will have no problems with it.

4.2.2. Kim

Figure 9: Portrait of Kim

Persona description

Kim is 38 years of age. He and his girlfriend, whose name is also Kim (!), have two kids, 3 and 5 years old. They live in a small yellow "radhus" in Slagsta, a working class suburb just south of Stockholm.

Since childhood, Kim has been playing with all kinds of technical equipment. At high school, he took the electrician's programme, and after that he did his military service in a tele-warfare unit. He never went on to higher education, since he doesn't like reading books.

Today, Kim works in a small company called MekanoData, together with 3 colleagues. His main task is to install and repair automatic ticket control machines at the Stockholm subway system.

There are two PCs at home. One of them is connected to the 30" flatscreen TV and surround stereo system and is used for watching downloaded movies and listening to mp3s. Sometimes the 5-year-old uses it to play games. The other (older) computer is used as an always on file server. Both run Mandriva Linux.

For work purposes, Kims girlfriend has a laptop computer running Windows XP. Kim would like for her to switch to Mandriva, too, but has not yet been able to persuade her. He likes the open source software philosophy, and argues that a switch could save in on anti-virus software cost.

Kim would like to be able to use the Mandriva Club as a help desk when running into problems with his system. He finds it quite difficult, but a few times he has found answers, posted questions and once he even answered a question by another user.

Goals

- Current goal: Have a cheap but well functioning computer system
- Long-term goal: Learn more, so that perhaps later he will be able to get jobs related to computer configuration and networking

4.3. Scenarios: a social semantic help desk at work

The scenarios below involve André, Kim and other users – Anna, Pedro, Renaud, Gonzo and Paul – who are shortly introduced in the respective scenarios.

4.3.1. André submits a question

André just bought a webcam and would like to know what software he can use for having video sessions with a friend of his. André logs on to the Mandriva Club, André writes his question in a simple text form: "what is the recommended software for video sessions?". The form also contains a set of optional statements for describing the context of André's computer work. These fields are pre-filled with what the Club knows about André's context, i.e. his hardware and software profile. André can either modify the field values, or check a box telling the fields are not to be taken into account. Once he has submitted his question, André gets a screen with following results:

- a list of resources from the Mandriva Club knowledge base related to the topics inferred from his question, with a summary of their content
- a list of similar questions that have received an answer, ordered by their social rank (see the scenario below on how to rank questions and answers)
- a list of experts available online for discussing the topic of video on Mandriva Linux live
- a list of external resources covering the topic, available on external knowledge bases

André notices that the first knowledge base resource displayed, titled "How to have video sessions with your friends?", is likely to bring him an answer. Hence he clicks the link and reads the article. He gets a list of software for having video sessions indeed, with some of them certified for the current version of Mandriva Linux he's using.

Contextually, the system displays a form asking André for feedback on how useful the resource chosen for solving his problem was.

4.3.2. André contributes with "how-to" information

André has bought a very recent drawing surface that he wants to connect to his Mandriva Linux computer. In doing so he runs into problems. While searching for help using the Mandriva Club search engine, he gets only links related to the use of that drawing surface with a Gentoo distribution. Looking into the way Gentoo community members managed to have the drawing surface work properly, he understands how he can draw inspiration from them for solving the problem in his context.

Since André has received a lot of help from Mandriva Club he wants to share his experience on the drawing surface with the community. André uses his recent search track on the Club and clicks on "Share experience". He clicks on "External devices" and "how-to", accepts the pre-chosen computer model and distribution he has, and enters the name and the model of the drawing surface. He describes the problem he had and how he solved it. This "how-to" contains information on what

the configuration file should look like for the drawing surface to work correctly.

4.3.3. Kim gives contribution feedback

Kim has been searching for a specific driver for his laptop and is presented with a link to information that André has contributed with earlier. Kim finds André's contribution very useful and clicks on the check-box saying that this information helped him solve the problem. When doing so, André's Karma increases, since he was the one contributing with the information. André can also see, on his Karma site, that Kim has successfully managed to solve a problem using his contribution.

4.3.4. Kim edits his personal notebook

Since Kim first became a Mandriva Linux user he has kept his personal notebook on his personal space in the Mandriva community. It helps him learn how to use, more efficiently, his system while putting into practice general how-to manuals.

By surfing around for information about his new laptop, he learned that in order to get his network card working he needs to change some parameters in the network configuration file. He finds this important to remember and clicks on the "My personal page" button on Mandriva Club. He uses the semantic editor on the site to make additional comments related to the network configuration resources available in the public knowledge base.

While saving the text, the system proposes an ontological representation of the text in order to make better searches for the information later on. This ontological representation is presented in natural language so that Kim does not need to know about the underlying RDF vocabulary. This representation consists of the following statements:

- this text is a personal comment authored by Kim on February the 25th, 2007, in English,
- it relates to following topics: network configuration,
- it relates to following products: Mandriva Linux 2007.0,
- it relates to following pages of the public knowledge base: Network_Configuration,
- it relates to following hardware type: Ethernet network card,
- it relates to following hardware model: 3COM-ETX-71661,
- it contains a configuration file example,
- the configuration file example was tested successfully with following hardware (here comes Kim's hardware profile),
- it relates to following RPMs: network-config.rpm.

Kim can then either accept this structured representation of his comment, or modify the statements using the semantic editor available online for changing the value of some properties. He can, for instance, add that he also successfully applied the script using the Mandriva Linux 2007.1 version as well.

As Kim also thinks that the notes he just entered might be of interest to others, he marks the text as public, thereby making the text available for others in the Mandriva community.

4.3.5. André searches for non Mandriva specific information in Linuxpedia

André currently has a problem with an installation script for his open source software. The script is working fine if he runs it from the hard disk, but not if he runs it from the CD, and he can't figure out why.

To solve the problem André enters Mandriva Club and searches for "installation script" and "open source software". He also checks the box "Linuxpedia" since he knows this is a general Linux issue. André gets three hits on the Linuxpedia page, among which two appear to provide him with the right information. He clicks on the links and together they provide enough information for André to make some clever conclusion about what it is he needs to do to get the installation script to work. André adds some comments to the Linuxpedia pages to help other Linux users solve the same problem.

4.3.6. Kim takes advantage of contextual help

Kim just bought a new camera from Logitech. While he plugs his camera into the USB port of his computer, an assistant offers him to browse the resources related to this camera available in the Club knowledge base. He has the possibility to:

- read tutorials in his language on how to take the most of this camera. He can choose to read tutorials that are available either in the community knowledge base or across his personal network of relationships knowledge bases
- read articles describing identified issues and workarounds related to the use of this given camera and the laptop model he has

4.3.7. Kim and Anna have a live help session

Kim's new Logitech camera is working fine with his laptop, except that the laptop emits a high pitch noise while the camera is plugged into the USB port. Kim submits the issue to the online help desk, but he cannot find interesting resources on the matter. However, a set of experts available for help on cameras is provided.

Kim asks one of them, Anna, for help. Anna is a member of the community who is working for Logitech and who has a long experience with dealing with Logitech cameras. Once she has read Kim's profile and the problem Kim is facing, Anna accepts to enter into a Live help session with Kim. Kim authorizes Anna to view his complete hardware and software profile. While perusing this configuration, Anna understands Kim has to update the driver of his video card. She brings this answer to Kim, with a link toward the new video card driver. The video card driver update fixes the problem indeed. Kim then reviews Anna's profile: among other expertise areas, Anna's profile indicates an expertise on cameras. Kim clicks the "thumbs up" button attached to that expertise. In the background, the help desk provides Anna with a summary in natural language of the context and message exchanged by both parties. Using a

semantic wiki editor, Anna can then enhance this summary, add statements to it so that the fix can be referenced immediately in the public knowledge base for others.

In doing so, while André two days later plugs in the same Logitech camera for the first time on his laptop, he gets notified by the help desk that he should update his video card driver for not hearing a high pitch noise when having the camera plugged into his computer.

4.3.8. A group of experts write collaboratively a manual on virtualization

As hardware virtualization technology use is spreading, more and more questions on the topic are issued by the user community (hardware virtualization is used for running more than one operating system at the same time). Within the Mandriva Club community, a set of ten experts having an in depth understanding of the topic and who have answered a high number of questions related to it decide to together write a manual entitled "Hardware virtualization using Mandriva Linux". For doing so, they create an expert group on the Club, and use the Nepomuk core services for establishing P2P connections between their desktops for working together on the topic during a period of three weeks.

The ten experts have accumulated on their desktop a large set of personal documents related to virtualization. André is part of the crew; he has 120 annotated documents on virtualization stored on his desktop, and a set of bookmarks, all together interlinked into a complex web of relations. This amounts to several hundreds of megabytes of data and meta-data. In order to share easily all their data in a common workspace, the team decides to use P2P communication, so that there is no need to transfer their large amount of data to a central place. Everyone can decide which part of his personal semantic web related to virtualization can be accessed by the others. Then this distributed workspace consisting of the aggregation of all individual semantic webs can be searched and annotated by all members of the group. The P2P communication mode is helpful for combining private semantic webs into a collective semantic web whose private entities can continue evolving as usual, on distinct desktops.

Group members can also edit some documents collaboratively in real-time. Using Nepomuk technology, they can work on the manual both while their connected to the other peers or not. When they get connected again, they can synchronize their personal web of data with the the updates brought by the rest of the team.

In order work in an efficient manner, the group first defines the global set of tasks that have to be performed when writing the manual, using the task manager component of Nepomuk. It lets them define specific and general activities, some of which can be directly reused from previous activity patterns identified by other groups while writing a manual.

The group extensively uses the existing question and answers issued by the community to make sure they neither forget key issues nor specific uses of the technology.

The manual is written using a semantic wiki editor, so that it is actually not only a plain text manual, but also a semantic graph of resources harnessing the Linux ontology.

Once the manual has reached a good quality level, it released to the Club community. The release is published in three main formats:

- a PDF file,
- a set of HTML static pages,
- a semantic graph of resources, that is published to the Club knowledge base as a named graph. The resources can then be enhanced, commented by the community, and new links between the statements of these resources and other statements can be drawn.

This semantic data can be harnessed by a semantic search engine and by an inference engine.

4.3.9. A group of experts write collaboratively a manual on virtualization

As Kim uses Mandriva Club help desk more and more, he gets to know a large number of users and experts, who progressively become good acquaintances that he trusts. Kim decides to create a group "friends" on the Club and adds his best online mates to the group.

In creating such a group of friends, Kim can narrow his search to the resources (knowledge base pages, experts, questions and answers, or web sites) that are rated as interesting resources by his network of friends. This feature lets Kim more efficiently discover new trustful experts and resources that may be of interest to him. While browsing available resources, Kim also gets the information on what his friends think about the resources he's browsing.

Using a personalized view of the Club home page, Kim decides to display a feed linking to the latest personal notes issued by his online friends: in getting to know his friends solve problems, Kim learns better and has a chance to discuss the related topics directly with people he gets on well with.

As time goes by, Kim creates other personal groups on the Club: one is the "family group", another one is the "colleagues" group, to which a few of Kim's work colleagues belong to.

4.3.10. Pedro using the Club help desk in combination with Linuxpedia

Pedro, a Gold member of the Club, is quite an advanced Linux user but even so, a new device in a new laptop model cannot not be his speciality. So this sound card doesn’t work with the latest Mandriva distribution.

Pedro has a "Contact Mandriva" button in his desktop, and being a Gold member he regards Mandriva as responsible for his problem. Clicking on "Contact Mandriva" opens a help desk new topic form, with his hardware data pre-filled. He notices that there is nothing about this novel sound card so he adds that information. He ticks a box saying that he’d like to see the post follow-up in his email as he’s not much on the web interface of the help desk.

As he submits, Pedro gets a list of a three Linuxpedia articles that seem to match his question. He is instructed to take a look at them. A Google search was also generated automatically from Pedro’s form post. The search opens in another browser frame.

Pedro checks the Linuxpedia articles and, comparing also with the Google info, thinks that one of them appears to match his problem, with some small differences, one difference being that the problem has only been spotted on Debian and RedHat. The Google search also confirmed another bug symptom that Pedro saw, so he edits the Linuxpedia article wiki-style. The article gets also marked automatically as "also seen on Mandriva". Pedro is happy and goes on to try the fixes that were tried on Debian and RedHat, but the fix on RedHat simply fails and for the one on Debian, the paths to the drivers in the file system are different. Pedro ticks some boxes in the Linuxpedia interface to mark that he tried the workarounds and failed.

Renaud is a Mandriva power user, who wants to get more karma and fancies with the goal of being paid part-time with Mandriva. When Pedro has confirmed that the sound card article is present for Mandriva, a bubble appeared on his screen. The bubble grew as Pedro marked the RedHat and Debian fixes as not working for Mandriva. Renaud clicks on the bubble and assumes from the log of actions taken that a possible new distribution-agnostic issue is on since this fancy new sound card appeared, and has not yet been confirmed on Mandriva but since it appears to be distribution-agnostic, it will be confirmed sooner or later. Renaud decides to take action and tries out the RedHat and Debian workarounds. He understands rapidly the difference in paths from Debian so he decided to try that first. Then he notices that there's no way that he can get the sound card work soon. He inserts the correct Mandriva path at Linuxpedia. That moment Pedro gets an email and an SMS (since he marked it urgent and allowed the use of SMS) that there is a possible fix on his problem. He tries out the Debian fix with the path entered by Renaud and things work. He marks the fix as working and Renaud's bubble starts to bounce, just as he was on his way out to buy crêpes. Upon lack of reaction from Renaud, the system also sends him an SMS, which she gets just after the crepes.

After devouring the crepes, Renaud goes back to his hacking spot enthusiastically and checks the Linuxpedia logs on the bubble. He now has a more complete picture of the issue and starts the steps necessary to introduce the fix in the next kernel versions. Renaud can't do it by himself as he doesn't have the sound card and the necessary rights, but he files a solution in detail, so he makes sure to get most of the karma on this issue. The solution can be found both by searching at Linuxpedia or by searching the Mandriva help desk.

4.3.11. Renaud is notified that many users are facing an issue with their sound card

Gonzo has got his Mandriva Linux from his father and just wants to use it to play games. He got this new laptop for his birthday, installed Mandriva and surprise, there's no sound in the game. Sound is very important to play the game properly, as the Bad Guys approaching can be heard and properly shot in due time. Upon losing some game points that he had work assiduously to accumulate, Gonzo gets anxious and slowly starts to think of doing something about this issue. He finally decides to use the "Contact Mandriva" button.

The help desk post opens, and he titles his post "the sound fails so i can't hear the Bad Guys" and gives some more detail but doesn't add any information about the sound card. When submitting, he is presented with five Linuxpedia links that correspond to sound card problems on his laptop type, and some Google search results. Gonzo doesn't have much

patience to read, so he clicks "I don't understand all this crap, contact me!" in the header. As more people like Gonzo fill in sound card problems for this laptop type, Renaud's bubble shivers more and more. Renaud has a strong suspicion that the problem can be fixed with this latest hack that he did the other day, for which he got a lot of positive feedback that increased his karma.

Renaud has lately been in close contact with Paul, a young Linuxer who, like Pedro, used to do a good job in spotting issues and suggesting fixes. Renaud decides to ask Paul to take care of the issue before it gets into the next kernel version, by communicating with the users who are not that skilled and have this issue. Renaud attaches the bubble into an email and sends it to Paul, mentioning also the karma that there still is to get.

4.4. Case study actors, data structures and processes

4.4.1. Case study actors

Several user types are involved in the case study. The Mandriva Club is comprised of members, some of which pay an annual fee for their membership while others don't. Paying members are categorized in 3 levels – Standard, Silver, Gold, that grant different access to Mandriva products, as summarized in the table below. Non paying members can be either VIP members, Mandriva employees, or Alumni. VIP members gain a VIP membership when they contribute in an outstanding manner to the Mandriva Linux project, either in coding, in testing the software or in translating documentation or software labels. Alumni are former members who did not renew their paying account.

Membership type	Description
VIP	VIP members are contributors to the Mandriva Linux project
Alumni	Alumni members are former members who have not renewed their membership. They keep having an access to the Club free services, mainly the Club forum.
Standard	Standard members can download the basic Mandriva Linux distribution including commercial drivers. They also have access to Mandriva eLearning service, and to the Club forum.
Silver	Silver members can download the advanced Mandriva Linux distribution. They have access to all Standard services.
Gold	Gold members can download most of Mandriva Linux software products.

Table 11: Different membership types on the Mandriva Club

4.4.2. Case study information and data

The core information involved in the case study will relate to the use of the Mandriva Linux operating system. Information will take the form of questions and answers in several languages, issued by community members. These questions and answers will be converted into first level how-to manuals or wiki pages facilitating reuse of information.

In addition to textual information, the data manipulated within the case study will consist of images or binary files that can be added to questions/answers, software code and binaries ranging from small

patches (size of a few kilo-octets) to large ISO files (5 Go), file signatures for checking the integrity of important files.

In terms of access rights, the system will have to deal with three types of contents:

Database	Description	Access policy
Public knowledge on Linux and open-source software	This database comprises public data related to Linux such as man pages, TLDP how-to ("TLDP" stands for The Linux Documentation Project), Wikipedia contents, etc.	The contents are open to all, both in writing and reading mode.
Mandriva Club contents	This database contains contents that are specific to Mandriva Linux system. Its include the documentation of the Mandriva Linux system, with many how-to manuals, question and answers produced either by Mandriva or by the Club community.	Most of these contents are open to all. Some of them are restricted to Club members.
Club members private databases	These databases live either on a single desktop, or across a set of desktop peers, who collaborate on the drafting on some documents and who want to keep the documents private. The documents may be released to the community at some point, but remains restricted to a set of collaborators during its elaboration	The contents are restricted to one user or to a set of peers.

Table 12: Case study databases description

4.4.3. Size of test material

As of September 2006, the Mandriva Club knowledge base consists of 800 documents. It is planned to considerably enhance the contents of the knowledge base during the case study lifespan, as well as the contents of a database dedicated to the description of open-source software use in general (i.e not distribution specific), by carrying out following tasks:

1. Importing data available from other existing databases,
2. Involving more largely the community of users and teaming up with other OSS content oriented projects,
3. Facilitating the publication and enhancement of contents through Nepomuk tools, and in particular through a rich client semantic wiki editor.

4.4.4. Case study main processes

The processes at the heart of the case study consists in submitting questions to the community, in bringing answers to open questions, in exploring the knowledge base graphically, and in rating available answers or experts. Beside these processes, the case study encompasses the production of how-to manuals based on the questions and answers.

Figure 10 illustrates the question/answer workflow.

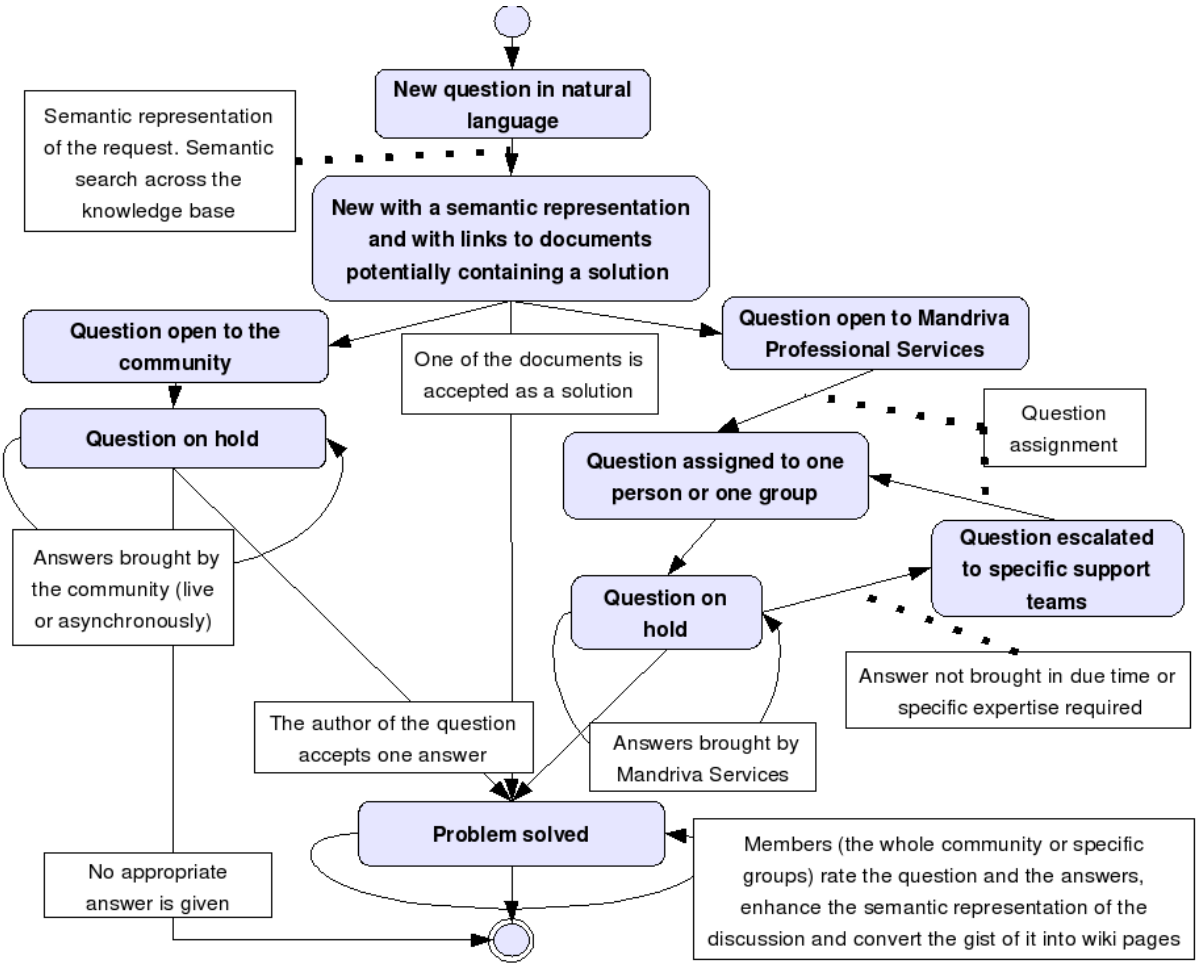

Figure 10: Help desk workflow

4.4.5. Data structures

The main data structures involved in the case study consist of classes in tables 13, 14, 15, and 16.

Data class: User	
Description	A user of the help desk (either expert or not)
Attributes	
Role	The role of the member: a member can be an expert, a moderator or a plain user.
Expertise areas	This field describes the list of expertise areas of the user if he is an expert.
Watch list	This field is used for notification purpose.
Configurations	This field stores the various system configurations of the user.
SIOC representation	This field stores the representation of the user in the SIOC ontology http://www.sioc-project.org.

Table 13: User data structure

A user can have one or several configurations, see table 14.

Data class: Configuration	
Description	A configuration consists of a software configuration and a hardware configuration. It describes a machine both from the software and the hardware point of view.
Attributes	
Software configuration	List of software installed on the machine, with their version number.
Hardware configuration	Precise description of the hardware components the machine comprises.

Table 14: Configuration data structure

Data class: Case	
Description	A case is a question.
Attributes	
Title	List of software installed on the machine, with their version number
Text	Precise description of the hardware components the machine comprises
Language	Language of the text
Open date	Date the case was created
Closed date	Date the case was set to "fixed"
Owner	Person who opened the case
Status	Current status of the case
Score information	Information related to the scores received by the community
Ontological representation	Representation of the case in the domain ontologies
Access level	The rule that applies for accessing or modifying the case, or for bringing an answer to it.

Table 15: Case class structure

Data class: CaseLog	
Description	A caselog is a message brought as an answer to solve the case, or to ask for clarification.
Attributes	
Parent Case	Identifier of the case this caselog refers to
Parent Case Log	Identifier of the caselog this caselog refers to (as answers can be brought to answers as well)
Language	Language used
Text	Body of the caselog
Date	Creation date of the caselog
Author	Author of the case log
Files	Potential attached files
Ontological representation	Representation of the caselog using the domain ontologies
Score information	Score data related to this caselog
Status change	Information related to the status change this case log may have brought to its parent case

Table 16: Case log class structure

4.5. Functional requirements

This section presents a set of functional requirements. For each requirement, the relevant Nepomuk workpackages beside WP11000 are mentioned.

4.5.1. Semantic wiki and knowledge workbench requirements

Requirement title	Wiki support for ontology design
Description	The wiki should have built-in support for collaborative ontology design. The wiki will be used both for entering data complying with one or several ontologies, and also for designing and maintaining the domain ontologies themselves. This means the wiki engine will have support for creating RDF classes and properties.
Workpackages	WP1000

Requirement title	RDF/OWL ontology import
Description	The semantic wiki engine should be able to import an existing RDF/OWL ontology and to create the corresponding wiki pages for each class and property of the ontology. The import should not overwrite systematically existing resources. It should rather ask the administrator of the system whether conflicting resources should be merged or replaced with the imported ones. See also http://wiki.ontoworld.org/wiki/Help:Import_vocabulary
Workpackages	WP1000 WP6000

Requirement title	RDF/OWL export
Description	The semantic wiki engine should be able to export the domain ontologies in RDF or OWL, as well as the entire semantic graph.
Workpackages	WP1000 WP6000

Requirement title	Support for semi-structured forms
Description	The wiki engine should support semi-structured forms. Semi-structured forms are forms into which both structured and unstructured data can be entered by the user.
Workpackages	WP1000

Requirement title	Ontology refactoring
Description	The ontology properties and types should be refactorable. The system should support the refactoring of the name and the properties of the managed resources. This means if the name of a resource is updated, all references to the resource should remain available in the system. The same remark applies to property name updates or property value changes. In this respect, the knowledge workbench should behave like an object oriented integrated development editor such as Eclipse that supports the refactoring of class or method names. The ontology evolution history should itself be stored so that one can understand the way the ontology was progressively designed. Implementation note: the backend should probably use an identification scheme for ontology elements so that name changes won't affect the whole corpus of documents.
Workpackages	WP1000

Requirement title	Embedded semantic documents
Description	The wiki editor should propose a syntax for embedding one or several semi-structured documents in another document. This means it will let the user create inner semantic documents. This is somehow similar to the support for anonymous class in the context of object oriented programming.
Workpackages	WP1000

Requirement title	Support for document combination
Description	The wiki engine should let users define documents comprised of other documents, or of other document parts.
Workpackages	WP1000

Requirement title	Questions issued in natural language can be converted to their representation in the ontology domain
Description	Questions submitted in natural language can be converted to their representation in the ontology domain, so that (i) proposed answers among those available are as relevant as possible, (ii) the experts suggested by the system as well as the experts notified of the question are as accurate as possible. The feature should work ideally with any language supported by the Mandriva Linux distribution (80 languages). The user then has the possibility either to enhance the proposed meta-data in adding or removing some properties, or to submit the question without reviewing the suggested ontological representation.
Workpackages	WP2000 (Text analysis)

Requirement title	Inline searches
Description	While typing some text, the user can issue an inline search query: the user can enter any SPARQL query anywhere in the text area and click a button to get the results displayed in the wiki editor itself. SemperWiki supports a similar feature. The requirement aims at helping experts answer questions and reduce the time needed for bringing accurate answers.
Workpackages	WP1000

Requirement title	Text completion based on the domain ontologies
Description	The semantic wiki editor should provide user assistance in proposing text completion while the user types some text, in a manner that is similar to what modern code editors propose.
Workpackages	WP1000

Requirement title	Semi-structured text comparison
Description	Users should be able to compare visually two versions of a document containing semi-structure data.
Workpackages	WP1000

Requirement title	Template system requirements
Description	The system should let users define advanced rules defining the way resources should be displayed, using templates. The rules will take into account the context. For instance a rule may say: use template "T1" for resources of type "How to", with some specific fields hidden when the user browsing the resource is a Linux newcomer.
Workpackages	WP1000 WP6000

Requirement title	Real-time collaborative editing
Description	Several users should be able to edit a common document in real-time, viewing what the other users are typing simultaneously, in a way similar to what is proposed by SubEthaEdit (http://www.codingmonkeys.de/subethaedit/).
Workpackages	WP1000

4.5.2. User interface

Requirement title	Zoomable semantic exploration of the help-desk question and answers and of the community knowledge-base
Description	Users should be able to explore the help-desk question and answers through a zoomable map representing clouds of terms and clouds of users. The users should be able to choose which information they want to see in the clouds: various semantic layers should be available. For instance, one should be able to show or hide the "hardware layer", the "software layer", the "user layer" etc.
Workpackages	WP1000

Requirement title	Visual refactoring
Description	Users with appropriate rights should be able to refactor the contents visually, changing their names, their status, their meta-data.
Workpackages	WP1000

Requirement title	Faceted browsing
Description	The system should let users explore the RDF graph using "faceted browsing" UI paradigm. This means the semantic graph can be explored from different points of view: topic centric, time centric, activity centric, project centric, association centric etc. See also Longwell browser description (http://simile.mit.edu/wiki2/Longwell) and Hyperscope
Workpackages	WP1000

4.5.3. Meta-data extraction

Requirement title	Data sources meta-data extraction
Description	WP11000 requires the extraction of meta-data from existing documents and databases available either publicly on the web (example: user manuals available on TLDP.org web site, or UNIX manual pages) or within Mandriva Club information system (example: Mandriva Club forum threads). The data sources for which the Semantic Data Access Framework should provide extractors are: wiki documents stored in a MySQL databases, PhpBB forum threads (plain text stored in a MySQL database), email archives stored by MhonArc archiving system, newsgroups archive, and desktop files in following formats: PDF, HTML, OpenOffice.org, mp3, mailbox, KDE data.
Workpackages	WP2000 (Semantic Data Access Framework)

4.5.4. Search

Requirement title	Support of semantic search
Description	Search within the help desk should harness the case study domain ontologies. The search engine will search across content resources, individuals or services. It will let users find experts matching a given set of characteristics, content material, or shared services.
Workpackages	WP6000

Requirement title	Search refinement along semantic and social axis
Description	The user must be able to refine the query results by restricting the values of some relevant properties and by narrowing the search using social criteria such as the quality level of the resources according to the scores given by the community, or by restricting the search to a given group's index.
Workpackages	WP6000

Requirement title	Search results ranking

Description	The user should be able to give a personal score to the results brought by the system. This score can be either private or shared with a group or the whole community of users. If the score is private, it will be used later on by the system for helping this specific user; if the score is shared, either with a given group or with the whole community, it can be used by the search engine for answering the queries of other persons.
Workpackages	WP5000

Requirement title	P2P search across a group of peers
Description	The system should let users search for some content across a set of peers using criteria related to the domain ontology. The distributed search engine should support the access rights management policy defined by the users.
Workpackages	WP4000

Requirement title	Social search support
Description	The search engine should propose personal recommendations based on the input from similar users.
Workpackages	WP5000

4.5.5. User context support

WP2000 tools detect the user context by observing which documents are touched by the user, as well as relevant information connected to the documents. The detected context information helps identify current processes and connects them to the workflow model. The context in which documents are used and manipulated is a key element in the enhanced local search infrastructure.

Requirement title	Ontological representation of desktop events
Description	When an event occurs on the desktop, such as a hardware related event (example: the user plugs a camera into the USB port of his laptop), the system should automatically generate an ontological representation of that event and propose the user to assist him in the context of his activity (for instance for searching for help on the installation or use of some new hardware).
Workpackages	WP2000

4.5.6. Knowledge work process support

Requirement title	Personal workflow support
Description	A service should provide a graphical assistant that lets the users define automated tasks he wants to run on this computer. These tasks should be fired either when a given event occurs (example: an expert with expertise in the area of webcams installation on Linux is now available online for real-time help) or periodically. The corresponding workflow to be triggered when the event occurs should be defined graphically by the user as well.
Workpackages	WP3000

Requirement title	Task pattern support
Description	Users should be able to define general task patterns for a given activity and to store them in the system for future use by them or others. Example: in the scenario above related to the collaborative writing of a manual, the involved users will be able to instantiate a "task pattern" describing their activity for coordinating their work, distribute the various tasks and continuously assess the progress toward the objective.
Workpackages	WP3000

Requirement title	Notification requirements
Description	Users should be able to express the notification they want to receive using advanced rules. Notification rules should support both content related events or user related events. Example: content updates, content creation, or user appearing on the network.
Workpackages	WP3000

Requirement title	Content synchronization
Description	Users should be able to modify offline some content that was fetched from an external database (external to the desktop). Users should then be able to synchronize their changes with those brought to the online document.
Workpackages	WP6000

Requirement title	Contextual recommendation
Description	While writing some document, the system should propose live related resources (documents or persons) that may be of interest to the user in the context of his current activity. Example: while an expert answers a question, he may request assistance from the system for getting directly while typing some relevant resources he will point the reader to.
Workpackages	WP2000, WP3000, WP5000, WP6000

4.5.7. Social networking

Requirement title	Group management
Description	The system should let users create groups and invite people to join groups.
Workpackages	WP5000, WP6000

Requirement title	Trust and reputation management
Description	The system should provide a mechanism for letting users assess the reputation of others so that a sound trust network can be set up between members.
Workpackages	WP5000

Requirement	Social network visualisation

title	
Description	The system should provide a way on each desktop to represent who is connected to the Nepomuk P2P network. Users can be displayed by group, by location, by topic or by any other property of the underlying social ontology.
Workpackages	WP5000

4.5.8. Desktop integration

Requirement title	Email integration
Description	It should be possible to ask questions, to submit answers, to receive answers and review answers by email.
Workpackages	WP2000

Requirement title	Desktop notifications
Description	The notification system should fire desktop events on the user end when a event he has subscribed to happens. Desktop events include the update of an applet in the user task bar, or the play of a sound.
Workpackages	WP2000

4.5.9. Access rights

Requirement title	Resource access rights management at the statement level
Description	The service responsible for managing access rights to resources should provide a mechanism for defining access right rules at the statement level of RDF resources.
Workpackages	WP6000

Requirement title	Service access rights management
Description	The service responsible for managing access rights to desktop services (such as P2P storage service, SPARQL query service etc.) should provide a mechanism for defining access right rules defining who can use which service.
Workpackages	WP6000

4.5.10. Openness and architecture for participation

Requirement title	Availability of an API for programmers
Description	The whole help desk key services should be scriptable and extensible. This means that the help desk contents and workflows should be accessible from external programs through an open API over the web, so that advanced users can create mashup applications harnessing the help desk API.
Workpackages	WP6000

4.5.11. System monitoring

Requirement title	Content statistics
Description	The system should compute periodically statistics on the evolution of contents available on the network. These statistics include the number of resources, the number of resource changes per period of time, the number of users connected, etc. The full list of metrics that the system should be able to compute periodically is available in the chapter below "Assessment criteria", in the paragraph pertaining to quantitative indicators.
Workpackages	WP6000

Requirement title	Network statistics
Description	The system should provide statistics on the P2P network, including: number of peers connected to the network, total bandwidth available.
Workpackages	WP4000, WP6000

4.5.12. Traditional content management requirements

Requirement title	General content management requirements
Description	The system should provide following standard ECM features (Enterprise Content Management): content versioning, conformance to content management standards (such as JCR) facilitating import/export of contents, content locking feature
Workpackages	WP6000

4.6. Non functional requirements

Requirement title	KDE, Eclipse, Firefox and Thunderbird integration
Description	The help desk should be well integrated with following desktop technologies: KDE, Eclipse, Firefox and Thunderbirds. This means a set of specific components harnessing the help desk capabilities should be available in the four frameworks.
Workpackages	WP6000, WP7000

5. Case study domain ontologies

Several domain ontologies will be designed and used in the context of the case study. These ontologies will be integrated with Nepomuk's mid- and upper-level ontologies.

5.1. Software use ontology

The software use ontology will describe the use of software on a computer, specifically a computer running a Linux operating system. It will provide a general framework for describing the activities of a user interplaying with a Linux computer and the issues that may occur. It will draw its inspiration from the following existing initiatives: Basic Linux Ontology, DebianTags, FLINK, EDOS.

5.1.1. FLINK – Formalizing Linux Knowledge

FLINK is a research project led by the Federal University of Bahia, Brazil[30]. FLINK stands for "Formalized Linux Knowledge". As stated on the web site of the project[31] the objective is "to demonstrate the benefits that will result from formalizing knowledge about the Linux operating system."

"Another way of phrasing our purpose is to say that we are assessing the benefits of Knowledge Management (KM) approaches to the Linux world. Specific questions of interest are:

- Are the issues facing the Linux world with respect to Knowledge Management any different from those facing other knowledge communities?
- What practical problems could KM tools and techniques solve for the Linux community? Are the standards emerging from the Semantic Web, RDF and OWL, of any practical relevance to the Linux community?
- Are there any other standards (Topic Maps, OpenCyc) better suited to this purpose?"

The FLINK team issued the following illustrative schema as an example of the basic elements of one of the targeted ontologies:

[30] http://www.dcc.ufba.br/

[31] http://flink.dcc.ufba.br/en/index.html

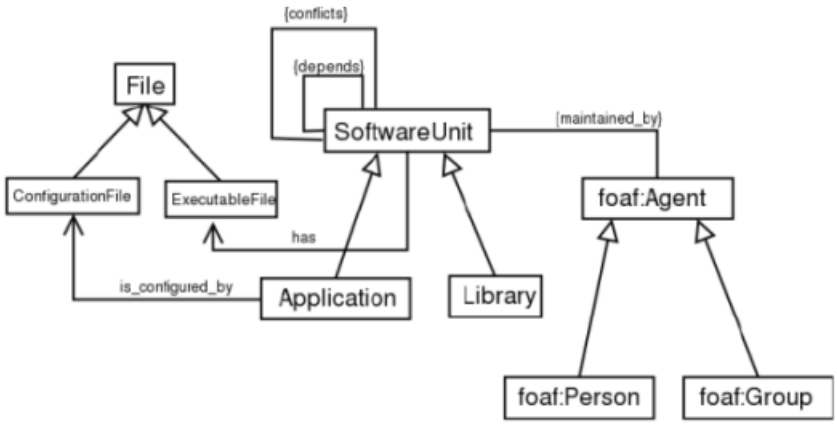

Figure 11: The basic elements of a package management ontology[32]

5.1.2. Basic Linux Ontology

The "Basic Linux Ontology" is an ontology available on the Internet at http://wwwis.win.tue.nl/~swale/blo . It contains a set of classes and properties describing a Linux operating system.

5.1.3. Debian Tags

DebianTags home page: http://debtags.alioth.debian.org/

DebianTags is a taxonomy providing a faceted classification of open-source software.

5.1.4. EDOS Project Management Interface

EDOS[33] is a Research Project funded by the European Commission as a STREP project. The project started October 2004 and ends June 2007. EDOS stands for Environment for the development and Distribution of Open Source software.

EDOS is designing a "Project Management Interface" described as follow: "The goal of EDOS Project Management Interface (PMI) is to define the key content and community artefacts of the F/OSS process and to formalize the relations between these. We believe that this precision allows inefficiencies in F/OSS processes to be detected and eliminated." The EDOS PMI initiative will be taken into account while designing the WP11000 domain ontologies. The TEAM EU Research Project[34] shares some objectives with Nepomuk, in the context of distributed software engineering: TEAM "addresses the need for a knowledge sharing

[32] http://www.ws.onto.ufal.br/Papers/sbia.pdf

[33] EDOS web site: http://www.edos-project.org

[34] http://team.iisa-innov.com/

environment with advanced capabilities suitable for the distributed engineering and management of software systems. The TEAM project aims to develop an open-source software system, seamlessly integrated in a software development environment for enabling decentralised, personalised and context-aware knowledge sharing through: Knowledge Desktop, Context Observer, History Analyser, Semantic Search, Semantic Recommendation, Metadata repository P2P Infrastructure". The project plans to produce ontologies. WP11000 team will keep abreast of the advancement of these ontologies.

5.1.4.1 TEAM EU Research Project

The TEAM EU Research Project[35] shares some objectives with Nepomuk, in the context of distributed software engineering: TEAM "addresses the need for a knowledge sharing environment with advanced capabilities suitable for the distributed engineering and management of software systems. The TEAM project aims to develop an open-source software system, seamlessly integrated in a software development environment for enabling decentralised, personalised and context-aware knowledge sharing through: Knowledge Desktop, Context Observer, History Analyser, Semantic Search, Semantic Recommendation, Metadata repository P2P Infrastructure". The project plans to produce ontologies. WP11000 team will keep abreast of the advancement of these ontologies.

5.2. Computer hardware ontology

The hardware ontology will provide a framework for describing the hardware pieces of a laptop or desktop computer. It will be directly based on the schema used by the HardwareLiSter utility, available for most Linux distributions.

HardwareLiSter[36] is described as "a small tool to provide detailed information on the hardware configuration of the machine. It can report exact memory configuration, firmware version, mainboard configuration, CPU version and speed, cache configuration, bus speed, etc. on DMI-capable x86 or EFI (IA-64) systems and on some PowerPC machines". "lshw displays nodes with attributes in a tree-like structure (that can be in indented plain text form, HTML, XML or graphically displayed in the GUI)."

5.3. User ontology

The user ontology will provide the framework for describing the social relations of the Mandriva Club community. The user ontology will be based on SIOC and FOAF.

---

[35] http://team.iisa-innov.com/

[36] http://ezix.org/project/wiki/HardwareLiSter

- SIOC: http://rdfs.org/sioc/ "SIOC (Semantically Interlinked Online Communities) Ontology provides main concepts and properties to describe the information from online community sites (e.g, bulletin boards, forums and weblogs) on the Semantic Web."
- FOAF: http://www.foaf-project.org/ "FOAF is about your place in the Web, and the Web's place in our world. FOAF is a simple technology that makes it easier to share and use information about people and their activities (eg. photos, calendars, weblogs), to transfer information between Web sites, and to automatically extend, merge and re-use it online."

5.4. Approach to work

By the end of 2006, the Nepomuk semantic wiki component will be put at the Mandriva Club community disposal within the Mandriva Club for designing collaboratively the domain ontologies of the case study.

The WP11000 team may get in touch with the initiatives mentioned above for potential collaboration in designing and maintaining the targeted ontologies.

6. Assessment criteria

The assessment task consists in measuring the evolution of the case study both quantitatively and qualitatively. This assessment will rely on indicators listed below.

As far as possible, the indicators listed below will be provided along several computing dimensions in addition to their absolute value at a given moment, i.e. per period of time, per group of users, per geographical localization, per resource language etc.

Assessments criteria used by KTH will focus on answering questions in a more qualitative way, through interviews, workshops and observing the user in action. In that respect, most of the indicators listed below are of secondary matter.

6.1. Content indicators

Content indicators
Number and size of documents available in the whole distributed system of peers
Number and size of documents produced
Number of questions submitted to the help-desk
Number of answers submitted to the help-desk
Percentage of questions finding an accepted answer

Table 17: Content indicators

Documents refer to files, wiki pages, resources having a "type", i.e. resources for which at least one statement with the predicate "rdf:type" exists in an RDF database of the system.

In the case of WP11000, the distributed system consists of:

- a common public database
- a database with some contents only available to Mandriva Club community
- a set of private contents. Measuring the size of private contents will depend on the users will to share some statistics.

6.2. Meta-data indicators

Meta-data indicators
Number of RDF statements available in the system in total or grouped by most relevant RDF types or properties
Number of RDF statements resulting from automatic processing (text-analysis, automatic tagging, etc.
Number of RDF statements entered manually by users
Number of statements created over a period of time

Table 18: Meta-data indicators

6.3. Social effects indicators

Social effects indicators
Number of groups formed
Sizes of groups
Percentage of accepted answers receiving reviews
Number of resource ratings produced
Number of person ratings produced

Table 19: Social effects indicators

6.4. System use indicators

Use indicators
Number of users using the system for producing contents
Number of users using the system for reading contents
Number of top level resources viewed per period of time
Number of top level resources produced per period of time
Number of search issued per period of time using the search form (distinguishing real searches from browsing by clicking tags or ontology elements that refine the search)
Number of users using exclusively the web based version of the system
Number of users using the rich client desktop version of the system
Number of new users registering to the service per period of time
Number of peers that are online in average

Table 20: Use indicators

6.5. Satisfaction indicators

The user satisfaction assessment will be conducted through a questionnaire comprising the questions below. The assessment will also be based on the analysis of the system with respect to the responses provided to a user requirements questionnaire.

Questions
On a descent scale from A to E, what grade would you give to the whole help-desk system?
How would you grade following modules of the system? Semantic editor module, search module, rating module, information visualization and browsing module, instant messaging communication module
Is it possible and easy for you to articulate your ideas and personal knowledge using the system?
Is it possible and easy for you to visualize your personal knowledge using the system?
How relevant are the resources recommended by the system?
Would you say you get answers to your question more efficiently than in 2006?
Which features are you missing?

Table 21: Satisfaction indicators

7. Conclusion

This study has shown how the Nepomuk components designed by the technical workpackages will be tailored to the Mandriva Linux community platform for bringing major improvements to the way information and processes are managed within the Mandriva Linux project.

The specificities of the case study – high number of participants, amounting to tens of thousands of users, very large domain ontologies covering software and hardware areas that evolve at a rapid pace, large amount of content material published along a various number of licenses (typically one of the Creative Commons licenses, with different levels of commercial friendliness) – will be addressed in adopting a very open and synergistic approach:

- the domain ontologies design process will follow the "folktology" approach instrumented by the semantic wiki designed by WP1000,
- the knowledge base component of the help desk will consist of two main parts: one, possibly in partnership with the EU SELF project, will consist of generic material on open-source software, while the other one will extend the latter in focusing on aspects that are specific to Mandriva Linux.

Next steps will consist both in progressively integrating Nepomuk components into the Club infrastructure and in developing specific components for the help desk. Major steps will consist in:

- the integration of a semantic wiki so that collective ontology drafting can start, in November 2006,
- the integration of a text analysis component for getting automatically semantic representation of help desk threads, easing the conversion of discussions into reusable knowledge,
- the integration of the trust framework that will let users assess the contributions and assess each other's expertise,
- the adoption of the P2P communication framework that will allow users to work collaboratively on common documents in private spaces.
- the use of a social search component that will bring community based recommendations to the users,
- the in depth integration of the help desk services with desktop technologies, that will increase the productivity of the participants.

The development methodology will follow the open-source software principle "release early, release often" and will continuously involve the end users.

In collaboration with all other Nepomuk workpackages, WP11000 aims at delivering one of the "killer applications" of the Social Semantic Web.

8. References

1. Nonaka, I. Takeuchi H. & Takeuchi, H. (1999) The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation.
2. Ontology is Overrated: Categories, Links, and Tags Clay Shirky http://www.shirky.com/writings/ontology_overrated.html
3. Creating Living Ontologies using Named Graphs, David Shotton and Chris Catton http://www.e-biosci.org/hinxton/Shotton.ppt
4. *Folktologies* -- Beyond the Folksonomy vs. Ontology Distinction[37], Nova Spivack http://novaspivack.typepad.com/nova_spivacks_weblog/2005/01/whats_after_fol.html
5. Construction And Evaluation Of OWL-DL Ontologies, Mark Wilkinson and Benjamin Good http://research.microsoft.com/workshops/FS2006/presentations/22_Wilkinson_071706.ppt
6. Sharing and Creating Knowledge in Open-Source Communities – The case of KDE, Andrea Hemetsberger , Christian Reinhardt http://www.belfin.at/marketing/content/download/ahcrpaper.pdf
7. Challenges of Online Reputation Mechanisms, Chrysanthos Dellarocas http://papers.ssrn.com/sol3/papers.cfm?abstract_id=393042
8. The Augmented Social Network: Building identity and trust into the next-generation Internet by Ken Jordan, Jan Hauser, and Steven Foster http://www.firstmonday.dk/issues/issue8_8/jordan/index.html
9. Earnings and Ratings at Google Answers, Benjamin Edelman Harvard University http://cyber.law.harvard.edu/people/edelman/pubs/GoogleAnswers-011404.pdf
10. Microsoft Social Computing Group http://research.microsoft.com/scg
11. IBM Social Computing Group http://researchweb.watson.ibm.com/SocialComputing/
12. IBM Watson Center Research Center - Collaborative User Experience (CUE) Research group http://domino.watson.ibm.com/cambridge/research.nsf/pages/cue.html?Open
13. MIT Viral Communications Lab http://dl.media.mit.edu/viral/
14. Howard Rheingold, Smart Mobs: The Next Social Revolution (2002)

[37] http://novaspivack.typepad.com/nova_spivacks_weblog/2005/01/whats_after_fol.html

15. Tim Berners Lee, Weaving the web
http://www.w3.org/People/Berners-Lee/Weaving/

16. Cooper, A. (1999). The Inmates are Running the Asylum: Why High-Tech Products Drive Us Crazy and How to Restore the Sanity. Indianapolis: SAMS.

17. Cooper, A. and Reimann, R. (2003). About face 2.0: The Essentials of Interaction Design. Wiley, Chichester, NY, USA.

18. Schrage, M, (1996) Cultures of Prototyping in Bringing Design to Software, Terry Winograd, 1996

19. Semantic Linux, a fertile ground for the Semantic Web, Guillaume Barreau and Debora Abdalla
http://www.ws.onto.ufal.br/Papers/sbia.pdf

20. Searching for a fresher taste, The Guardian
http://technology.guardian.co.uk/weekly/story/0,,1667276,00.html

21. Fast cheap zero cost curation model for ontology building, Benjamin M. Good, Erin M. Tranfield, Poh C. Tan, Marlene Shehata, Gurpreet K.Singhera, John Gosselink, Elena B. Okon, and Mark D. Wilkinson
http://helix-web.stanford.edu/psb06/good.pdf

22. Harvesting Wiki Consensus - Using Wikipedia Entries as Ontology Elements, Martin Hepp, Daniel Bachlechner, Katharina Siorpaes
http://www.heppnetz.de/files/SemWiki2006-Harvesting%20Wiki%20Consensus-LNCS-final.pdf

23. Roadmap for Tool Support for Collaborative Ontology Engineering, Yiling Lu
http://www.cs.uvic.ca/~chisel/thesis/YilingLu.pdf

24. Tom Gruber Interview in AIS SIGSEMIS Bulletin Oct.2004
http://www.sigsemis.org/newsletter/october2004/tom_gruber_interview_sigsemis+collective+ontology+engineering

25. Motivating participation by displaying the value of contribution, Al M. Rashid, Kimberly Ling, Regina D. Tassone, Paul Resnick, Ann Arbor, Robert Kraut, John Riedl
http://portal.acm.org/citation.cfm?id=1124772.1124915

26. Social Search: Bringing the Social Component to the Web. Gerd Stumme, University of Kassel, Germany.
http://research.microsoft.com/ur/us/fundingopps/RFPs/Search_2006_RFP_Awards.aspx#Social_Search:_Bringing_the_Social_Component_to_the_Web

27. Knova help desk datasheet
http://www.kanisa.com/docs/datasheets/Knova-for-Help-Desks.pdf

9. Glossary

Folksonomy	A folksonomy is an Internet-based information retrieval methodology consisting of collaboratively generated, open-ended labels that categorize content such as Web pages, online photographs, and Web links. A folksonomy is most notably contrasted from a taxonomy in that the authors of the labeling system are often the main users (and sometimes originators) of the content to which the labels are applied. The labels are commonly known as tags and the labeling process is called tagging. Source: Wikipedia http://en.wikipedia.org/wiki/Folksonomy
Folktology	A folktology is the combination of a folksonomy and an ontology: "in a folktology, users could instantly propose or modify ontological classes and properties in the same manner that they do with tags in tagging systems." Source: *Folktologies* -- Beyond the Folksonomy vs. Ontology Distinction[38], Nova Spivack
Helpdesk	A help desk is an information and assistance resource that troubleshoots problems with computers and similar products. Corporations often provide help desk support to their customers via a toll-free number, website and/or email. There are also in-house help desks geared toward providing the same kind of help for employees only. A typical help desk has several functions. It provides the users a central point to receive help on various computer issues. The help desk typically manages its requests via help desk software that allows them to track user requests with a unique ticket number. This can also be called a "Local Bug Tracker" or LBT. The help desk software can often be an extremely beneficial tool when used to find, analyze, and eliminate common problems in an organization's computing environment. The user notifies the help desk of his or her issue, and the help desk issues a ticket that has details of the problem. If the first level is able to solve the issue, the ticket is closed and updated with documentation of the solution to allow other help desk technicians to reference. If the issue needs to be escalated, it will be dispatched to a second level. Source: Wikipedia http://en.wikipedia.org/wiki/Help_desk
Incident escalation	Incident escalation is the act of advancing an issue to the next appropriate level for resolution.
Incident Management	Incident Management (ITSM) is an IT Service Management process area. The first goal of the incident management process is to restore a normal service operation as quickly as possible and to minimize the impact on business operations, thus ensuring that the best possible levels of service quality and availability are maintained. 'Normal service operation' is defined here as service operation within Service Level Agreement (SLA). ITIL terminology defines an incident as:

[38] http://novaspivack.typepad.com/nova_spivacks_weblog/2005/01/whats_after_fol.html

Any event which is not part of the standard operation of a service and which causes, or may cause, an interruption to, or a reduction in, the quality of that service

Incidents may match with existing 'Known Problems' (without a known root cause) or 'Known Errors' (with a root cause) under the control of Problem Management and registered in the Known Error Database (KeDB). Where existing work-arounds have been developed, it is suggested that accessing these will allow the Service Desk to provide a quick first-line fix. Where an incident is not the result of a Known Problem or Known Error, it may either be an isolated or individual occurrences or may (once the initial issue has been addressed) require that Problem Management become involved, possibly resulting in a new problem record being raised.

The main incident management processes are the

- Incident detection and recording
- Classification and initial support
- Investigation and diagnosis
- Resolution and recovery
- Incident closure
- Incident ownership, monitoring, tracking and communication

Source: Wikipedia
http://en.wikipedia.org/wiki/Incident_Management_%28ITSM%29

ITIL

The Information Technology Infrastructure Library (ITIL) is a framework of best practices approaches intended to facilitate the delivery of high quality information technology (IT) services. ITIL outlines an extensive set of management procedures that are intended to support businesses in achieving both quality and value for money in IT operations. These procedures are supplier independent and have been developed to provide guidance across the breadth of IT infrastructure, development, and operations.

Source: Wikipedia http://en.wikipedia.org/wiki/ITIL

Karma

Karma is a "Sanskrit word meaning action, effect, destiny. Karma is a sum of all that an individual has done, is currently doing and will do." (Source: http://en.wikipedia.org/wiki/Karma). The term was introduced into the world of online communities by the Slashdot platform (http://www.slashdot.org), which replaces the real-world word-of-mouth mechanisms with a reputation scores in order to provide incentives to all members to contribute positively to the community. The score of each individual is dubbed the "karma".

MVP

Microsoft Most Valuable Professional: The Microsoft Most Valuable Professional (MVP) Program is an award and recognition program run by Microsoft. Microsoft MVPs are volunteers who have been awarded for providing technical expertise towards communities supporting Microsoft products or technologies. An MVP is awarded for contributions over the past year.

Source: Wikipedia. http://en.wikipedia.org/wiki/Microsoft_MVP

PEBKAC

PEBKAC is an acronym which stands for "Problem Exists Between Keyboard And Chair". Also used is PEBCAC, which stands for "Problem Exists Between Computer And Chair", or PBKAC, which stands for "Problem Between Keyboard And Chair". Sightings of PEBCAK ("Problem Exists Between Chair And Keyboard") have been reported. Another

	variation is PIBKAC ("Problem Is Between Keyboard And Chair"). It is most commonly used by experts to describe to one another that the problem was not in the computer but was instead caused by the user operating it. Source: http://en.wikipedia.org/wiki/PEBKAC
Reputation system	"Reputation system collects, distributes, and aggregates feedback about participants' past behavior" Source: P. Resnick, R. Zeckhauser, E. Friedman, E., and K. Kuwabara, 2000. "Reputation Systems: Facilitating Trusted Internet Transactions," Communications of the ACM, volume 43, number 12, pp. 45-48.