

Case Study: KDE 4.0 Semantic Desktop Search and Tagging

Leo Sauermann, DFKI GmbH, and Sebastian Trüg, Mandriva Linux

May 2008



Introduction

KDE (K Desktop Environment) is a [Free Software](#) environment for desktops and portable computers that has been created by an international technology team. Among KDE's products are a modern desktop system (window manager, file manager, desktop search), comprehensive office productivity and groupware suites, and hundreds of applications for Internet and web, multimedia, entertainment, education, graphics, and software development. KDE software is translated into more than 60 languages and is built with ease of use and modern accessibility principles in mind.

KDE version 4.0 is a significant release that has been available since January 11, 2008. The libraries, desktop, applications, and artwork have been improved and adapted, and it now runs natively on a wider set of operating systems that include Linux, BSD, Solaris, Windows and Mac OS X. As part of this release, the search engine was overhauled and desktop annotation introduced. Based on standards and technologies developed in the [NEPOMUK Semantic Desktop](#) project, the search engine is now based on RDF and allows users to annotate and search for their files using Semantic Web standards.

You can download and use [KDE 4.0](#) yourself today.

Requirements for KDE4

The K Desktop Environment provides a platform on which a vast amount of applications are built, including Konqueror for file browsing, and KOffice for office applications. An open community for applications exists through [KDE-Apps.org](#), where thousands of applications are available. In total, it is estimated that there are more than a million KDE users.

It was decided for KDE4 that a capable desktop search engine was needed with functionality for fast full text search, and semantic search for enhanced retrieval accuracy. Support for tagging and annotations of files was also requested by many users. As KDE is a platform, it was important that the search and annotation services were made available **across all applications**, allowing users to annotate and find documents independent of the program or file format.

In addition to these functional requirements, KDE software has to conform to standards, demonstrate strong performance, and be stable. New capabilities have to be documented well so that developers from the application community can easily build or extend their applications. As a rule of thumb, a developer should not need to take more than 30 minutes to understand the library and start to use it.

The Solution for Search and Annotation

For the new KDE version 4.0, the [NEPOMUK EU project](#) contributed a semantic search and annotation engine called [Soprano](#), based on RDF. The desktop search service provides an integrated search on all information the user has stored on their computer. This allows users to search for e-mails, appointments, contacts, and files using a single interface. The NEPOMUK project was able to provide such facilities by exploiting metadata in powerful ways. The core elements of NEPOMUK have already been integrated into KDE.

Using RDF, Soprano allows developers to index and search documents independent of application format, file format, or other implementation requirements. Given that thousands of new file and application formats may be added by the various applications that KDE supports, the **extensibility** of RDF was considered a key feature.

How NEPOMUK-KDE Works

NEPOMUK provides functionality for handling all of the metadata on the KDE desktop in a generic fashion. The three sources of the metadata include:

- Metadata that is automatically captured in files by applications, e.g. timestamps, or simple text strings.
- Metadata created manually by the user, e.g., a comment assigned to a file.
- Metadata that is automatically captured about transient resources, e.g. the URL of a Web site that was viewed, or a document that was an attachment to an email.

NEPOMUK allows users to manually tag, rate, and comment on files using the KDE file manager (see [Figure 1](#)). The KDE file analyzer can then be used to index file metadata, so that it becomes possible to view files according to related metadata attribute, rather than just through placement in a file system (see [Figure 2](#)). In KDE applications, users are able to annotate files using tags (such as "work", "to do", "technology", "project x") and use these labels to retrieve similar files. Based on this technology, personal information management is much simplified, as it becomes possible to retrieve entities based upon subject matter, rather than file type or storage location (see [Figure 3](#)).

This capability is provided through the use of Semantic Web standards. RDF is used as the mechanism for assigning tags to files because it is powerful and extensible. RDF Schema (RDFS) is used to create a simple ontology that is used for normalizing the tag terms, so that all files of interest can be retrieved.

Future development plans include embedding search functions into more KDE applications, thereby providing powerful query capabilities regardless of the application that is being used. A particularly interesting use case is allowing users to perform arbitrarily complex semantic queries to retrieve emails through a selection of virtual folders in KMail. For example, to *show mails from people who I have met in meetings in the last week*.

User Experience

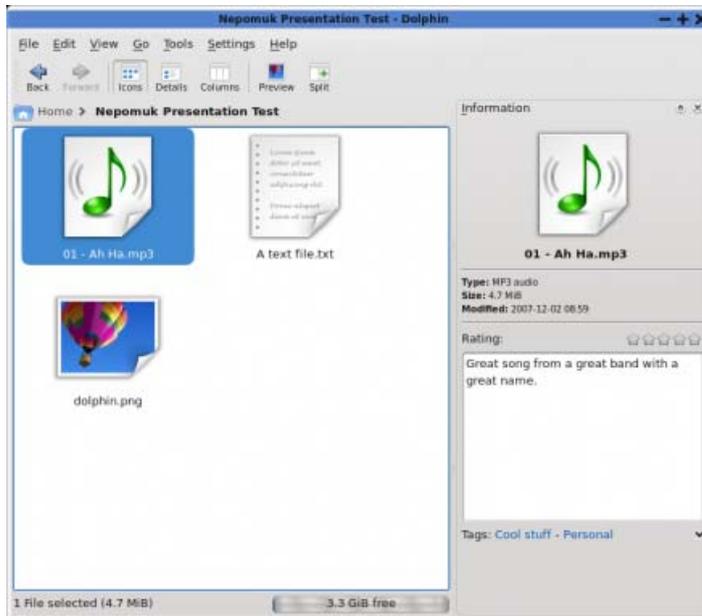


Figure 1: The KDE File Manager for allowing users to tag, rate, and annotate files (a larger version of the image is also available).

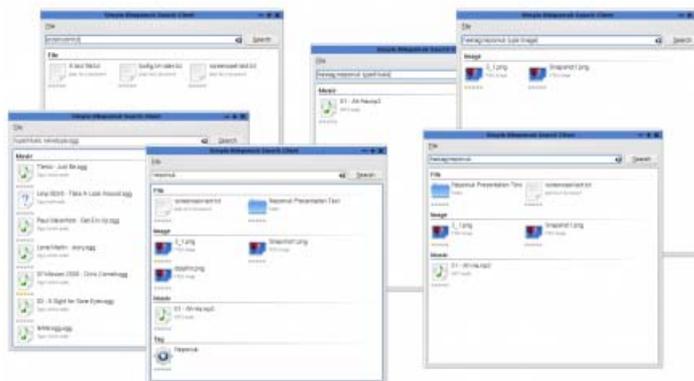


Figure 2: The KDE File Analyzer can be used to index metadata, thus enabling users to perform queries across file metadata, and tags and rating systems (a larger version of the image is also available).



Figure 3: Interface for users to add ideas, projects, todos or any other item of interest in an extensible data architecture, the *Personal Information Model (PIMO)*. For example, a concept of “friend” can be used to add friends to the system.

Key Benefits of Using Semantic Web Technology

- RDF is a standardized format that is documented, patent-free and royalty-free. For an open-source project, using standardized technology has the benefit of avoiding legal problems and makes developer uptake easier.
- Semantic Web standards allow data integration across desktop applications.
- The standards enable search across multiple desktop applications.
- RDF gives users the freedom to add more annotations (tagging, ratings, personal comments, free text).
- The RDF data model is very extensible, which is important when it is possible that the range of supported data formats will need to be extended by KDE or individual developers.
- RDF provides a clean and powerful way to expose the natural object-oriented organization of data on the desktop. Data can be interlinked and organized in whatever way the user desires.

Key Problems of Using Semantic Web Technology

At the time of implementing, using OWL and rules to infer new knowledge decreased performance considerably: no fast and open source OWL inference engine was available. The need was for an extensible and adaptive data model and fast storage, so we decided to limit the system on using RDF/RDFS.

Another problem was developer uptake due to the lack of documentation of RDF at the beginning of the project. Given the large community of KDE, a very long phase of teaching, discussions, and evangelizing the idea of RDF was needed. Once the key benefits of RDF were understood by the stakeholders involved, support for the project increased.

Summary

The implementation of the search capability within KDE 4.0 was based on RDF and RDFS. The key features of interest with these technologies included the extensible and adaptive data model, and fast storage and retrieval of data.