

Miipa-Doc ANR Project



Multi-dimensional indexing with a tagging application leading to and using faceted classification

About me

- Last year PhD student in Information Sciences
- In the school engineering Conservatoire National des Arts et Métiers (CNAM) France (Paris)
- Research Laboratory : DICEN (Information and Communication System in Digital Age)
- Director : M. Manuel Zacklad
- Thesis Subject : The combination of heterogeneous KOS, stakes and perspectives
- Taking part in ANR Project Miipa-Doc

Presentation of Miipa-Doc project

- ◉ Integrated institutional and participative methodology and services for facet based classification for complex documentary contents
- ◉ Include the development of :
 - > A software
 - > A methodology
- ◉ Partner : EDF R&D (nuclear and thermal energy producer)
- ◉ 11 persons Team

Research fields explored

- ◉ Ergonomics
- ◉ Occupational psychology
- ◉ Sociology of organizations
- ◉ LIS
- ◉ Communication
- ◉ Cognitive sciences
- ◉ Computer sciences

Context

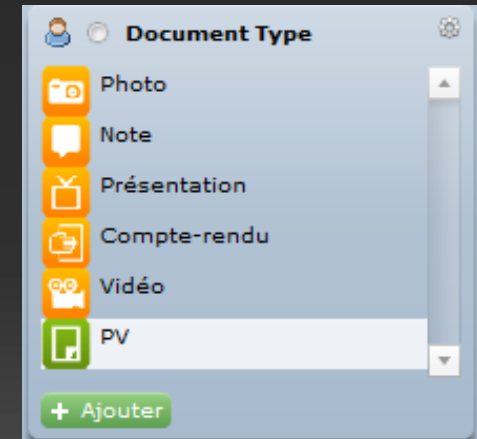
- Focus on working documents : all digital documents produced, used, individually or collectively in the everyday working life
- Using search engines (generally in full text) = general answer to information management problems in organizations
- Search engines should be used in combination with others KOS

Objectives of the project

- Facilitate « documentarisation » (Zacklad, 2005)
 - Consists in adding metadata (or attributes) to documents for a future use in perspective of conservation of communication transactions associated to the documents
- Facilitate searching for documents during working activities
- Add perenial metadatas to documents
- Provide a multi-dimensional approach to indexing and research

Semiotag as new KOS

- Each tag associated with an icon
- Groups of tags with similar nature => facets
- Each Facet can be considered as a taxonomy
- Groups of facets are called contexts
- Contexts can represent
 - > activities
 - Study
 - Design
 - Construction
 - Test
 - > or other organizational elements
 - Account services
 - HR service
 - Direction service
 - ...



=> For indexing and finding documents

Software functionalities

- ◉ Individual Tagging
- ◉ Collaborative tagging for sharing documents : sharing tags and facets
- ◉ Indexing : creating title for the document and adding metadata
- ◉ Classification of the documents and management of storage issues
- ◉ Rules : combinations of tags can become rules for classification and storage
- ◉ Feeding on the fly of the facet classification
- ◉ Finder : filtering with tags included in facets to find documents
- ◉ Interoperability with Electronic Documents Management systems

Benefits

- ◉ Ergonomic man-agent interface
- ◉ Flexibility
- ◉ Simplified Management of facet classification
- ◉ Multi- dimensional indexing
- ◉ Fast Finding documents with facet filtering
- ◉ Possibility to avoid using arborescences' organization

My contribution

- Conception of the methodology accompanying the software
- Study and propositions about the main concepts of the project: facets, tags, icones, « Semiotags », facet classification, context
- Tests
- Functionnal specifications conception

Purpose of the methodology

- Purpose : find the best scenario for the deployment of the software in
 - > Small organizations
 - > Medium organizations
 - > Large organizations
- In a perspective of a progressive approach to constraints
 - > First scenario : free
 - > Second scenario : moderately restrictive approach
 - > Third scenario : restrictive approach
- Degree of constraint are not necessarily correlated to the size of the organization

Priority to the medium scenario

- ◉ Priority to bottom up approach
- ◉ With contribution from the users (Feeding on the fly of the facet classification)
- ◉ Enhanced gradually
- ◉ Needs of a light study for the beginning
- ◉ To suggest a basis for the facet classification
- ◉ Reach a compromise between free and restrictive approach

Crossing study of

- Professional activities (specific)
- Information organization activities
- Concrete organization of informations (including the KOS)
- Types of groups needing to share documents
- Document types created, used, archived, related to work activities

Applications of the methodology

- How to transform an existing KOS into faceted classification?
 - > Arborescences (individual or shared)
 - > Folksonomies
 - > Databases
- Others controlled languages as resource
 - > Thesaurus
 - > Ontologies

Other approaches

- ◉ Restrictive approach:
 - > The study is improved (more precise)
 - > For purposes like
 - Archiving
 - Corporate information management
- ◉ Free approach
 - > For Personal Information Management
 - > On personal computer
 - > For individual activities
 - > Done by the user

Observations

- For one professional activity (including different tasks) one can be in situation of needing the 3 scenarios
 - > One can be free to organize his resources
 - > But has to share part of the documents he produces
 - > And has to leave some of them available for the whole organization

Perspectives

- We need to test the methodology with
 - > Other users than our partners (in process)
 - > For different uses
 - Archiving
 - Sharing documents
 - Individual self-sufficient users for their individual activities
- We need to test the software with a scaling up number of users