Search User Interface proposal for Subject Repositories: DSpace implementation for WindMusic.org

June 21st 2011 - OAI7 DSpace User Group Meeting Université de Genève (CH)

Christophe Dupriez, dupriez@destin.be

Table of Content

Context	1
WindMusic.org	2
Resources discovery	2
Thesaurus and authority lists: controlled vocabularies	3
Query Reformulation patterns	4
WindMusic implementation	4
browse categories or indexes to discover the overall content and structure of database	
formulate a query: specialized search boxes with menus or terms auto- completion	6
3. remove the necessity of reformulating queries to take into account synonyms, translations and generic versus specific concepts (the search for the general cainclude its specifics)	
4. browse search query results looking without disruption to individual record to discover typical documents, possible authors, terminology, abstracts	8
5. pick interesting resources for later in depth review	.10
6.1 easily add criteria to make a query more specific	11
Adding multiple criteria in a single step	.13
6.2. easily remove any search criteria	.14
Bibliography	.14

Context

Subject repositories are targeting exhaustivity within a domain. They want to be a central source of information for researchers: they tend to be very large. Those repositories also require precise indexing and controlled multilingual vocabulary for efficient retrieval of documents answering the explicit and tacit interests of the researcher.

After presenting in Göteborg (DSUG 2009, Sweden) how we integrate thesauri and authority lists (organized along the W3C SKOS standard) into DSpace, we present to DSUG 2011 what discovery tools are now available for the users of WindMusic.org.

We hope that this presentation will be felt very practical and that it will spring ideas and discussions for potential evolutions of DSpace.

WindMusic.org

The CDMC (Conseil Départemental de Musique et de Culture de Haute Alsace = Upper Alsace Council for Music and Culture) is an organization supported by public founding since 1969. It promotes the cultural production (music, dance and theatre) of the Département of Upper Alsace. The CIM (CDMC Music Information Centre) provides the trainees, teachers, musicians and conductors a **structured information tool**_targeted to the needs of Wind Orchestras and Bands: WindMusic.org.

Many types of resources are available: from encyclopedias to CD-ROMs, with periodicals, monographs, but mainly scores (with about 34.000 items) and audio recordings (with about 30.000 items). WindMusic is multilingual: French, German and English.

This computer based catalog allows visitors to discover full scores of interesting works and, in the CDMC library, to listen directly their recordings with headphones on multimedia stations (Audio recordings are only available within the CIM to respect publishers' copyright).

WindMusic was been re-implemented in 2008 using DSpace and a whole set of functions to manage and take advantage of SKOS standard (authority lists and thesauri).

Working for the CDMC and for the Belgium Poison Centre, DESTIN provided the development services necessary for this realization. These results are there to share in projects with the DSpace community.

Resources discovery

Searching within a domain is a process where a person takes an authoritative source, gives some clues about her/his needs and learns about the related database content before digging further. Formulating and reformulating queries, interesting resources are hopefully gathered.

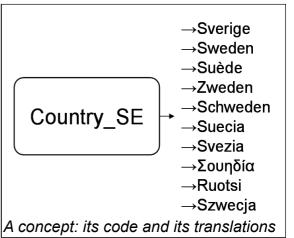
Tools are therefore necessary to:

- 1. browse categories or indexes (possible search criteria) to discover the overall content and structure of the database,
- 2. formulate a query:
 - 1. "à la Google" search box,
 - 2. using specialized search boxes with menus or terms auto-completion.
- 3. search taking into accounts synonyms, translations, generic versus specific concepts (the search for the general can include its specifics). This removes the necessity of reformulating queries. Controlled vocabularies, automated query expansion...
- 4. browse search results looking (without disruption) to individual record to discover typical documents, possible authors, terminology, abstracts...
- 5. pick interesting resources for later in depth review:
 - 1. add / remove resources from the pick list,
 - 2. save, print or export the pick list in different bibliographical references formats.
- 6. improve a query with easy to make reformulations:
 - 1. add criteria to make the query more specific,
 - 2. remove criteria to make the query more general.

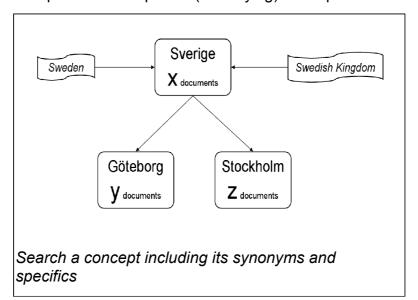
Thesaurus and authority lists: controlled vocabularies...

On the road to build this collection of tools, we already tackled many issues related to terminology (SKOS thesauri addition to DSpace presented at DSUG 2009):

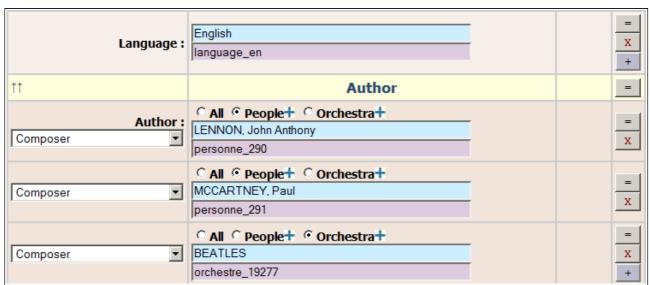
- indexation is based on "concepts", each concept being identified as a code stored in the database;
- concepts can be retrieved using their names in different languages or synonyms;
- menus or auto-completion input boxes assist the user when choosing a concept;
- concepts' usage statistics are gathered and displayed to help users discriminate between frequent and rare concepts;



 concepts can be put in hierarchies going from generic to specific: retrieval may include a concept and all its specific (underlying) concepts:



 The metadata update form has also been overhauled to take advantage of controlled vocabularies:



Query Reformulation patterns

Below, we use the DSpace syntax for queries (implied AND):

- ABC means: term-A AND term-B AND term-C
- A (B OR C) means: term-A AND (term-B OR term-C)

Liu and Gwizdka in "Analysis of query reformulation types on different search tasks" have estimated the frequency of the different ways a user may reformulate her/his search queries. From their results, we can draw the following table:

Liu& Gwidzka Code	Reformulating Query A B C into Queries like:	Frequ- ency	Operation
G	AB, BC or AC	15%	Delete some criteria
GR	AX, BX or CX	12%	Delete some criteria, add some but less
S	ABCX	20%	Add some criteria
SR	ABXY, BCXY or ACXY	12%	Replace some criteria and add some more
WR	ABX, AXC or XBC	19%	Replace some criteria
REPEAT	ABC, ACB, BCA, BAC, CAB or CBA	4%	Query unchanged
R2	Reformulation using synonyms	5%	
N	Unrelated new query	13%	
		100%	

- Unrelated new queries are only 13%: the support for incremental query formulation is essential.
- Indexing expansion with Thesauri Synonyms relieves users from about 5% of the reformulation needs.
- Adding criteria is necessary in GR, S, SR and WR: 63% of the reformulations (80% of the non trivial cases). Faceted browsing of search results is a well accepted tool for this.

When multiple criteria are added at once, it remains to be studied what are the relative frequencies for:

```
1. A B C (X OR Y OR Z)
2. A B C X Y Z
```

- 3. other combinations of AND, OR, NOT Boolean operators
- Deleting criteria is necessary in **G, GR, SR** and **WR**: 58% of the reformulations (74% of the non trivial cases). Little buttons to delete specific parts of a query is a straightforward way to do this.

We tried to adapt the DSpace search user interface to fulfill these needs.

WindMusic implementation

We implemented different tools for the main functionalities identified for Resources Discovery:

1. browse categories or indexes to discover the overall content and structure of the database

Main categories:

Communities and Collections

Shown below is a list of communities and the collections and s name to view that community or collection home page.

1. Wind Music Scores

- 1.1. Concert Band Scores [18931]
- 1.2. "Batterie-Fanfare" Scores [1176]
- <u>1.3. Big Band Scores</u> [947]
- 1.4. Other Scores (Fanfare Band, Brass Band) [608]

2. Wind Music Recordings

- ■2.1. Concert Band Recordings [25548]
- ■2.2. "Batterie-Fanfare" Recordings [1611]
- 42.3. Big Band Recordings [549]
- 42.4. Other Recordings [2264]

3. Ensembles

- 3.1. Scores [1429]
- 43.2. Recordings [836]

4. Solo and Educational Works

- <u>1</u>4.1. Solo Works [10165]
- 4.2. Educational Works [1172]

5. Other Resources

- \$\sum_{5.1}\$. Books [1346]
- 95.2. CD [3482]
- \$\mathbb{O}\$5.3. Audio-visual (Videotapes, DVD) [12]
- 5.4. CD-ROM [13]

6. Around Musical Works

- 6.1. Authors (Biographies) [19630]
- 6.2. Publishers [1950]
- \$\infty\$6.3. Orchestra [1664]
- ■6.4. Keywords (index) [1019]
- 6.5. Series [1634]

(number of linked resources)

Subject Index:

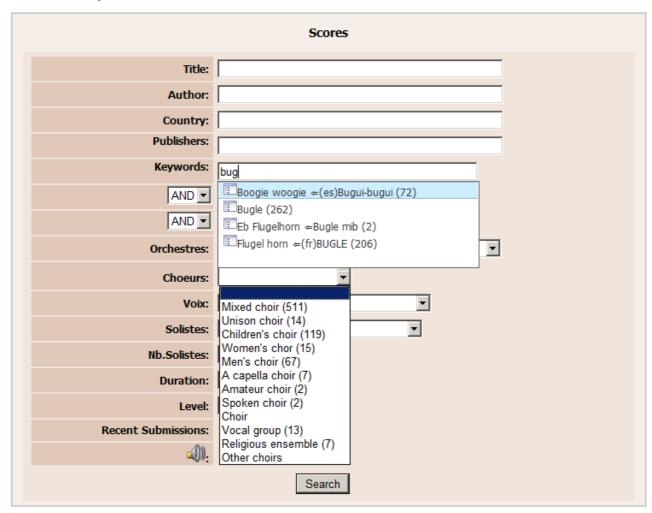
- Music (61866)
 - □ Ensemble (32914)
 - □ Epoch (1179)
 - Music playing (206)
 - Music style (16559)
 - Musical creation (175)
 - Musical events (89)

 - Musical genre (26533)
 - Musical writing (54686)
 - Score (1439)
- Music instruments (16667)
 - □ Electroacoustic (28)
 - Keyboards (739)
 - Percussions (1951)
 - Strings (3277)
 - Voice (722)
 - □ Winds (11394)
- Performing arts (5519)
 - ☐ Cinema (83)
 - □ Circus (206)
 - □ DANSE (5126)
 - ☐ Incidental music (8)

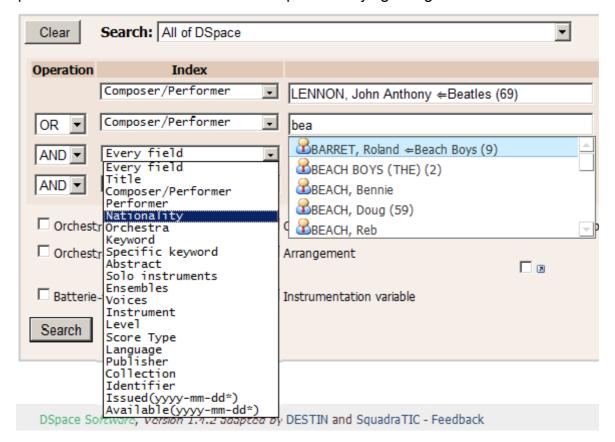
 - Musical-theatre (10)
 - Performing arts (67)
 - □ Puppet theatre (1)
 - □ Theatre (132)
- Period (56176)
 - 11th century (2)
 - □ 12th century (11)
 - □ 13th century (9)
 - □ 14th century (12)
 - □ 15th century (18)
 - □ 16th century (258)

 - □ 17th century (624)
 - □ 18th century (1814)
 - □ 19th century (4217)
 - 20th century (33461)
 - □ 21st century (16212)

2. formulate a query: specialized search boxes with menus or terms autocompletion..

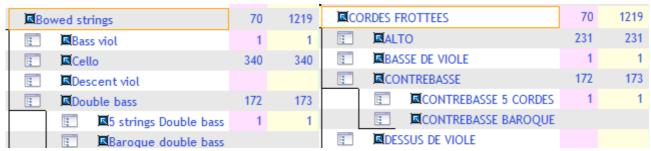


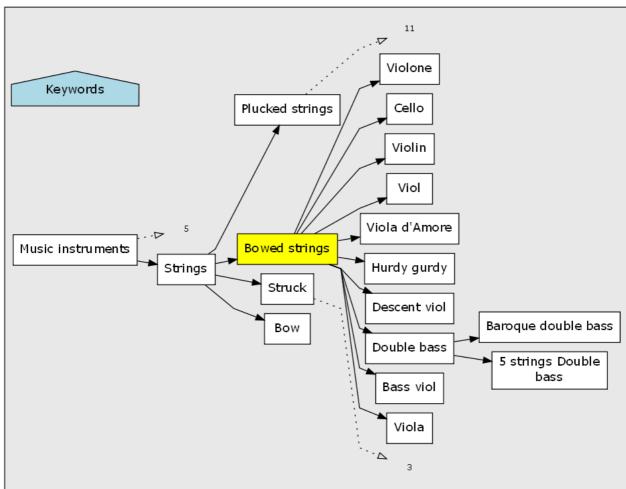
DSpace Advanced Search with auto-completion varying along chosen index field:



3. remove the necessity of reformulating queries to take into account synonyms, translations and generic versus specific concepts (the search for the general can include its specifics)

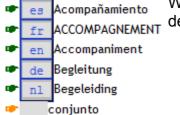
"Bowed Strings" is mentioned in 70 references but, with its specifics, it is linked to 1219 references:





Concept sujet_21963 (displayed using ASKOSI.org)

Spanish, French, English, German and Dutch translations:



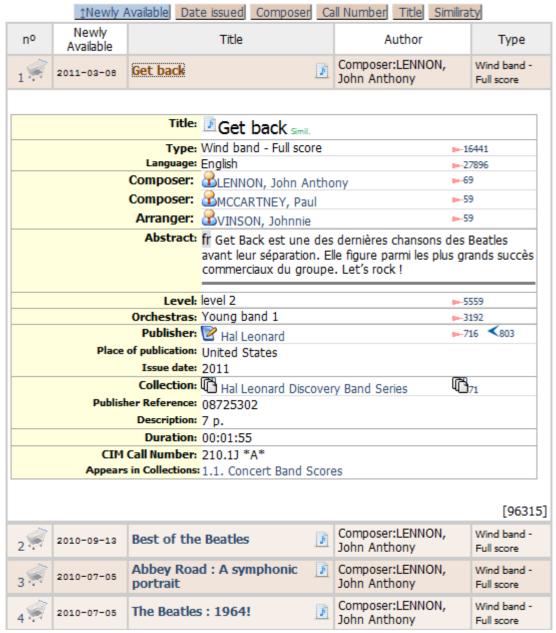
Word search can be done using any synonym or any translation defined for a concept in the SKOS thesaurus.

4. browse search query results looking <u>without disruption</u> to individual record to discover typical documents, possible authors, terminology, abstracts...

Clicking the title of resource within a search result:

author: & LENNON, John Anthony OR author: & BARRET, Roland x

Results 1-78 of 78.



Item display is presented within the search result

Clicking on another title:

author: BARRET, Roland_x OR author: BARRET, Roland_x

Results 1-78 of 78. Newly Available Date issued Composer Call Number Title Similiraty Newly nº Title Author Type Available Composer:LENNON. Wind band -Get back 2011-03-08 1 3 John Anthony Full score Composer:LENNON, Wind band -Best of the Beatles 2010-09-13 John Anthony Full score Title: Best of the Beatles smil. Additional Title: fr Le meilleur des Beatles smil. Contents list: Ticket to Ride - Hey Jude - Get back Type: Wind band - Full score -16441 Language: English **-27896** Composer: LENNON, John Anthony **69** Composer: MMCCARTNEY, Paul -59 Arranger: Moss, John -81 -1 Level: level 2 **-5559** Orchestras: Concert band 1 42575 Additional instruments: Glockenspiel 1 13 sol. 13 sol. Note about instrumentation: Instrumentation variable à 5 voix et percussions. Publisher: W Hal Leonard **-716 ≺803** Place of publication: United States Issue date: 2010 Collection: Flex-Band Series (P)23 Publisher Reference: 04002952 Description: 19 p. **Duration:** 00:04:00 CIM Call Number: 210.1V *A* Appears in Collections: 1.1. Concert Band Scores [94154] Abbey Road : A symphonic Composer:LENNON, Wind band -2010-07-05 portrait John Anthony Full score

• Clicking on a title, the user receives the record display inside the current search result. Clicking the title again hides the record display.

Composer:LENNON,

John Anthony

Wind band -

Full score

 Clicking on any other title, the previous record is hidden and the new one is inserted.

The Beatles: 1964!

2010-07-05

• Database maintainers can still click on the number at the left of each line to display the record in a separate web page.

This removes most disruptions when inspecting the records within a search result.

5. pick interesting resources for later in depth review

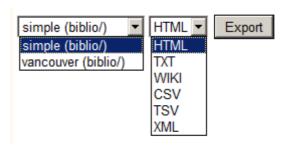
Items 2 and 4 has been clicked and added to the "caddy" :: :

author: BLENNON, John Anthony OR author: BARRET, Roland



These two items can be:

- removed from the caddy (clicking again on each item caddy or canceling the whole caddy content)
- 🖶 printed
- put in a separate search result for:
 - sorting: <u>Newly Available</u> <u>Date issued</u> <u>Composer</u> <u>Call Number</u> <u>Title</u>
 - export in different bibliographic format (HTML, Wiki, Text)
 - · batch update using Excel.



6.1 easily add criteria to make a query more specific

(Liu & Gwidzka codes GR, S, SR and WR: 80% of the non trivial cases)

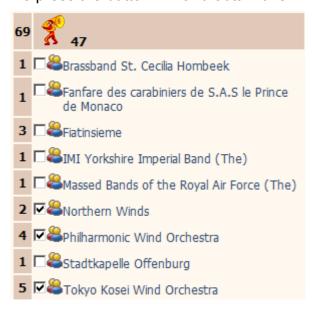
Faceting panel is added to every search result. It sorts all the controlled vocabulary used by the different records in the result and it offers different options to create a new query.

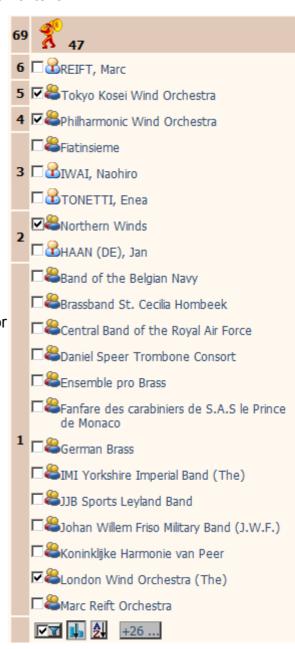
In WindMusic, faceting is done along many search criteria:

- 1. Composers
- 2. Performers (same authority lists than Composer, different role)
- 3. Documents types
- 4. Keywords
- 5. Publishers
- 6. Difficulty levels for the performer
- 7. Instruments
- 8. Solo instruments (same vocabulary, different role)
- 9. Collections, serials
- 10. Languages

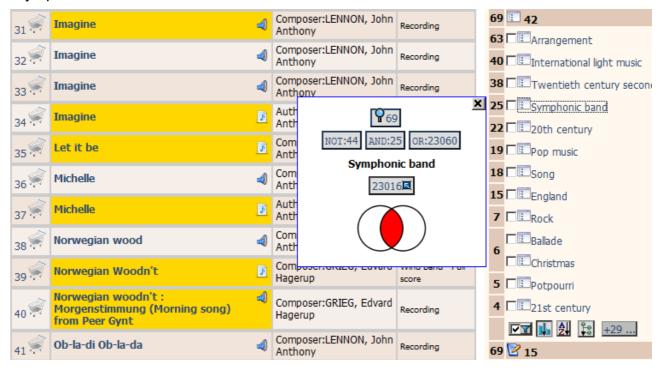
In this example, we only see the faceting done for performers. If we only wish to examine records done by four orchestras we prefer, we can select them (5 records by Tokyo Kosei Wind Orchestra, 4 by the Philharmonic Wind Orchestra, 2 by the Northern Winds and 1 by The London Wind Orchestra) and then click the funnel button.

If we prefer to see the list in alphabetical order, we press the button and obtain this:

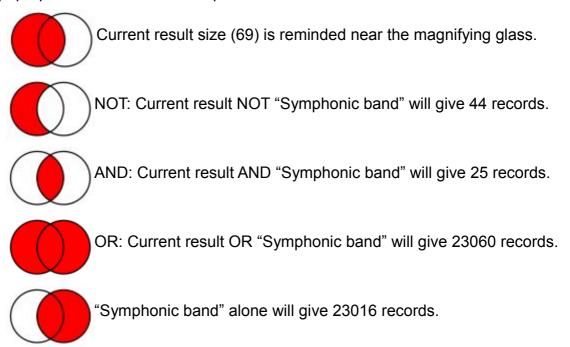




Any concept listed in the faceting panel can be clicked. For instance, with Keyword "Symphonic band":



- 1. Records in the result for a "Symphonic band" are immediately highlighted: you can look at them, they will be displayed in place, no disruption.
- 2. A pop-up menu offers Boolean operations:



3. Just click another concept to explore other possible results...

Adding multiple criteria in a single step

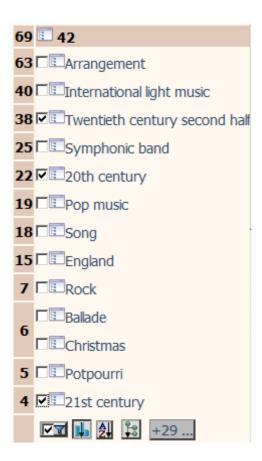
It was shown above how multiple orchestras could be selected. The resulting query is:

composer: LENNON, John Anthony_x AND performer: London Wind Orchestra (The)_x OR Tokyo Kosei Wind Orchestra_x OR Philharmonic Wind Orchestra_x OR Northern Winds_x)

Results 1-12 of 12.

This shows that adding criteria "X Y Z" (all from the same facet) to a query "A B C" creates a query: A B C (X OR Y OR Z)

We could have selected the orchestras but also some keywords (for instance, keywords indicating the century):



This generates a longer query:

author: LENNON, John Anthonyx AND keyword: 21st centuryx OR London Wind Orchestra (The)x OR Tokyo Kosei Wind Orchestra OR Philharmonic Wind Orchestrax OR Northern Windsx

This shows that adding criteria "X1 Y1 Z1" and "X2 Y2 Z2" (from two different facets) creates a query: A B C (X1 OR Y1 OR Z1) (X2 OR Y2 OR Z2)

- Logical AND between the previous query and between different facets,
- Logical OR between the concepts within a same facet.

6.2. easily remove any search criteria

(Liu & Gwidzka codes **G, GR, SR** and **WR**: 74% of the non trivial cases)

The search query is presented like this:

author: LENNON, John Anthony OR author: BARRET, Roland

Clicking the little red "x" after *LENNON*, *John Anthony* removes that criteria from the query. The new result is:

Résultats 1 à 9 sur un total de 9.

	1Nou\	veauté Date de publication	Compositeur Cote Titre	Ressemblance
no	Nouveauté	Titre	Auteur	Туре
1	2010-03-07	Final impact	Compositeur:BARRET, Roland	Orchestre à vent - Orchestration Complète
2	2009-02-16	St. Thomas	Compositeur:ROLLINS, Sonny	Orchestre à vent - Orchestration Complète
3	2009-02-16	Night Train	-	Orchestre à vent - Orchestration Complète
4	2009-02-16	Of Dark Lords and Ancient Kings	Auteur:BARRET, Roland	Orchestre à vent - Orchestration Complète
5	2009-02-16	Festival Sinfonia	Auteur:BARRET, Roland	Orchestre à vent - Orchestration Complète
6	2009-01-05	Of Dark Lords and 4 Ancient Kings	Compositeur:BARRET, Roland	Oeuvre sur CD
7	2009-01-05	On the Winds of the d Chosen	Compositeur:BARRET, Roland	Oeuvre sur CD
8	2009-01-05	Of Dark Lords and 4 Ancient Kings	Compositeur:BARRET, Roland	Oeuvre sur CD
9	2009-01-05	Fanfare and Dance Segments	Compositeur:BARRET, Roland	Orchestre à vent - Orchestration Complète

Bibliography

- Analysis of query reformulation types on different search tasks. Liu, Chang; Gwizdka, Jacek. iConference 2010 poster. https://www.ideals.illinois.edu/handle/2142/15049
- 2. **Search Patterns: Design for Discovery**. Morville, Peter; Callender, Jeffery; O'Reilly Media 2010. http://oreilly.com/catalog/9780596802288
- 3. **Search User Interfaces**. Hearst, Marti. Cambridge University Press 2009. http://searchuserinterfaces.com
- 4. WindMusic, example of the new possibilities for DSpace when adding SKOS thesaurus and authority lists management. Dupriez, Christophe; Schubnel, Julien. DSUG 2009. http://hdl.handle.net/2077/21341