

1. Introducció

La Biblioteca de la Universitat de Girona és l'encarregada de coordinar els repositoris digitals de la Universitat. A l'inici el repositori neix de la necessitat de difondre i preservar tota la recerca de la Universitat, però mica en mica el projecte va creixent i es creen diferents repositoris per a diferents tipus de continguts. Així es creen els repositoris DUGiDocs, DUGiMedia i DUGiFonsEspecials.

El DUGiDocs conté els documents digitals de la recerca i la docència generada pels professors de la Universitat. El DUGiMedia conté els àudios i els vídeos dels congressos, actes i conferències de la UdG. I finalment el DUGiFonsEspecials conté les digitalitzacions dels Fons Especials de la Biblioteca de la UdG.

A més a més, la Biblioteca participa en els següents repositoris consorciats: TDX, MDX, MDC, RACO, Recercat, Recolecta, DRIVER i OPENAIRE. Els primers cinc repositoris estan gestionats per el CBUC, que és el consorci de biblioteques de les universitats catalanes. El TDX és el repositori que conté les Tesis lligades a les Universitats Catalanes. L'MDX és el repositori que agrupa tot el material docent de les universitats catalanes. L'MDC conté les digitalitzacions dels fons especials de les Biblioteques Catalanes. El RACO recull les revistes catalanes. I el Recercat conté la recerca generada a les Universitats Catalanes. Els altres tres repositoris són d'àmbit espanyol i europeu, i recol·lecten dels repositoris de les Universitats.

Amb tota aquesta diversitat de repositoris es crea la necessitat d'un punt d'entrada comú a tot aquest material. D'aquesta necessitat neix el DUGi 1.0, que pretén ser el portal d'entrada i cerca conjunta a tots els material de la UdG.

Pel DUGi 1.0 es va utilitzar un DSpace, al qual se li va afegir un plugin que era un client d'OAI-PMH, un servei que permet recol·lectar el contingut dels repositoris, utilitzant un script de Perl es recol·lectaven els repositoris i s'importaven al DUGi 1.0 utilitzant el seu procés d'importació d'ítems. Per a recol·lectar el RACO es necessitava un script de Perl que parseges el RACO, ja que aquest no tenia servidor d'OAI-PMH, i l'importés al Dspace fent servir el mateix procés de DSpace.

2. Problemàtica

El DUGi 1.0 no era una solució prou bona, ja que tenia els següents problemes:

- El client d'OAI no entén els atributs d'alguns tags d'XML.
- El client d'OAI no detecta més d'un tag d'XML en una mateixa línia.
- Sobren moltes característiques del DSpace.
- El DSpace consumeix masses recursos.
- Necessitem acoblar masses tecnologies.
- Es perd el qualificador de les metadades.
- La interfície no és clara ni usable.
- No existeix un control d'autoritats.
- Les opcions de cerca no ens convencen.

Amb tots aquests problemes, s'entreveu la necessitat de crear una aplicació que els resolgui i aporti una solució més elegant i precisa. D'aquí sorges el DUGi 2.0 que es presenta en aquest PFC.

3. Objectius

L'objectiu d'aquest PFC és desenvolupar un portal web amb un sistema d'indexació que permeti la cerca per diferents criteris com, per exemple, autor, matèria i títol.

L'entorn ha de tenir un disseny agradable per a l'usuari final i ha de permetre obtenir les dades del repositori que se li indiquin. També ha de ser capaç de detectar els possibles duplicats en els criteris de cerca i notificar-ho a l'administrador.

Per tal d'aconseguir aquest objectiu, ens hem proposat les següents tasques:

- Fer un estudi de les eines més interessants per implementar aquest projecte.
- Trobar una metodologia que ens pugui servir per dissenyar i implementar el portal d'una manera eficient.
- Dissenyar l'aplicació
- Implementar-la
- Fer la instal·lació tant de l'entorn de treball com de l'entorn de producció
- Fer el testeig
- Fer una optimització
- Fer un estudi de les principals amenaces de seguretat que ens poden afectar, i en cas d'haver-n'hi corregir-les.

4. Requisits

Veient les problemàtiques del sistema anterior ens en donem compte que necessitem un sistema ideat especialment per a l'ocasió. Per això els requisits que hem definit per aquest projecte són:

- Crear un sistema que ens permeti unificar les entrades d'autors per evitar les confusions en els resultats.
- Implementar un client d'OAI-PMH completament estàndard i que ens permeti aprofitar totes les opcions del protocol.
- Enriquir les cerques donant opcions de filtre i refinament.
- Proporcionar un protocol de cerca prou potent per anar de la cerca més senzilla a la més complexa.
- Simplificar al màxim la interfície per aconseguir una alta usabilitat.
- Donar la màxima visibilitat als autors, matèries i documents.
- Implementar un sistema de memòria cau per fer les cerques més ràpides.
- Ha de tenir una interfície multilingüe. Com a mínim ha de donar suport per als idiomes català, espanyol i anglès.
- Ha de ser segur.
- Ha de consumir pocs recursos.
- Ha d'estar format de mòduls reutilitzables.

5. Desenvolupament del projecte

Per desenvolupar el projecte s'ha dut a terme les següents tasques:

- Anàlisi dels requisits del sistema

En aquest punt es van definir totalment totes les funcionalitats que havia de tenir el sistema, i es van fer els diagrames de casos d'ús i les fitxes de casos d'ús.

- Estudi de les tecnologies i eines necessàries

Aquest estudi va ser molt productiu perquè ens va permetre conèixer i provar eines molt interessants. Finalment vam escollir un sistema de plantilles per a PHP i el servidor de cerca SOLR.

- Disseny conceptual

En aquest punt s'han fet els diagrames de classes necessaris per definir el model de dades del sistema.

- Disseny navegacional

En aquesta etapa vam fer el disseny de la navegació, amb les relacions entre les pàgines, utilitzant la metodologia UWE, que és la metodologia UML adaptada per al desenvolupament d'aplicacions web.

- Disseny de la presentació

En aquesta altra es va fer tant els diagrames UWE de presentació com el prototip de la interfície del sistema.

- Instal·lació de la màquina

En aquest punt vam instal·lar tots els serveis i llibreries necessaris sobre una màquina virtual bàsica amb el sistema operatiu Ubuntu Server Edition.

- Implementació

La implementació es va haver de fer en menys temps del que s'havia previst, entre 6 i 7 setmanes, perquè es va necessitar prioritzar altres tasques. El sistema es va implementar de forma modular i aplicant alguns patrons de disseny per fer un codi més reutilitzable.

- Testeig

L'aplicació es va testejar amb l'ajuda de varis dels integrants del departament. Així es va poder detectar petits errors de funcionament i funcionalitats que cal millorar en versions futures.

- Correccions

Paral·lelament amb el testeig es van anar fent les correccions.

6. Conclusions

Aquest ha estat un projecte enriquidor del que hem après a utilitzar i a pensar amb bases de dades NOSQL, en aquest cas més concretament amb l'índex de cerca SOLR. També hem utilitzat una llibreria de plantilles per PHP amb un resultat més que acceptable, i hem fet servir el format YAML per als fitxers de traducció i de configuració de la recollida de cada repositori.

Finalment ha quedat una aplicació complerta i amb un acabat força professional gràcies a l'ús de mòduls d'apache com per exemple *mod_deflate*, *mod_rewrite* i de bones pràctiques de seguretat. A més a més, s'ha optimitzat l'aplicació per a que els cercadors puguin indexar el contingut i donar el màxim de visibilitat possible als autors i als ítems continguts al DUGi.

També estem molt satisfets perquè els objectius proposats prèviament s'han superat perfectament, i els requeriments també s'han complert quasi al 100%.

Malauradament, tot i que els resultats han estat bons, hi ha hagut alguns punts que cal millorar en pròximes versions, ja que finalment tot i que hem creat una sistema de cerca molt potent, no hem pogut crear una interfície de filtre de resultats tant usable com esperàvem.

Igualment faltaria permetre el canvi del nombre de resultats per pàgina per fer usable el llistat de resultats.

També creiem que es poden millorar els sistemes d'emmagatzematge de les estadístiques i les autoritats per a comprovar utilitzant el sistema de bases de dades REDIS, que s'ha estudiat a la fase d'estudi de les tecnologies i eines necessàries.

Per últim faltaria acabar d'implementar l'accés directe als documents compresos als ítems, lligant la visualització dels vídeos amb el servidor d'streaming, i oferint una interfície per consultar els documents de text sense haver-los de descarregar.