



A DiVA rendszer bemutatása és a további fejlesztések

Háttér

2000-ben az *Uppsalai Egyetem* elhatározta, hogy projektet indít az egyetem valamennyi publikációjának elektronikus megjelenítésére. Az *Elektronikus Publikációs Központ (Electronic Publishing Centre = EPC)* 2000 szeptembere óta működik. Célja, hogy publikációs lehetőséget nyújtson az egyetem oktatóinak és hallgatóinak, a tudományos eredmények megjelenítésének és archiválásának könnyen kezelhető eszköze legyen. A tárhelyen elhelyezett digitális dokumentumok strukturáltan kereshetők a könyvtári katalógussal és más releváns szolgáltatással egyidejűleg. Metaadataikat különböző információs rendszerek hasznosíthatják.

Erre a projektre épült a *DiVA (Digitala Vetenskapliga Arkivet = online tudományos archívum)* rendszer, amely nyílt szabványokon és - ahol lehetséges volt - nyílt forráskódú szoftvereken alapul. A metaadatokhoz és a teljes szövegű fájlokhoz az Unicode-ot és az XML-t választották. A fejlesztési időszakban a szövegfájlokhoz a PDF-et használták, kezdetektől fogva XML támogatással.

A DiVA az egyetem publikációs rendszeréből nemzetközi együttműködéssé nőtt, 17 egyetem részvételével Norvégiából, Svédországból és Dániából. 2002-ben a DiVA-t használó egyetemek konzorciumot alakítottak az eszközök, a tapasztalatok és a költségek megosztására.

A rendszer több, Java és XML technológiára épülő modulból áll. Felépítésének köszönhetően további modulok is bekapcsolhatók.

A rendszer bemutatása

A DiVA 2003 januárjában indult a következő modulokkal: tárhely, publikáló rendszer, szerzői modul, indító portál, kereső modul, terjesztési modul, archívum. További modulokat is fejlesztenek majd: a publikációt megelőző munkák folyamatvezérlője, az archiválási folyamat vezérlője.

A DiVA *tárhelye* teljes szövegű doktori téziseket, diplomamunkák téziseit, kutatási jelentéseket és

cikkeket, valamint metaadatokat tartalmaz. A repozitórium közvetlenül a fájlrendszerben helyezi el a fájlokat. A fájlrendszert független szerveren helyezték el, és bármelyik modulból távolról is elérhető egy titkos kód használatával. Hozzáférést csak az a modul kap, amelynek kódját a tárhely felismeri.

A DiVA *publikáló modulja* a nyomtatást és az elektronikus publikálást egyaránt támogatja. A metaadatok közzétételére több lehetőség is van: például webes interfész, ahol a szerző a dokumentumot és a metaadatokat együttesen töltheti fel, de igénybe vehet MS Word űrlapot is. A metaadatok használhatók a dokumentumok címlapjának, a könyvek borítójának és weblapjainak elkészítéséhez. A publikálás után a rendszer a dokumentum adatait automatikusan elküldi *Stockholmba*, a *Svéd Nemzeti Könyvtárban* lévő *Nemzeti Katalógusba*.

Fejlesztik még a teljes szövegű fájlok XML-ben történő megjelenítését szolgáló modult, amely lehetővé teszi majd, hogy az XML fájl bármilyen szövegszerkesztő stíluslapjának felhasználásával megjelenjen. A tervezett munkafolyamatban a beérkező dokumentumokat az *Open Office* segítségével nyílt dokumentumformátumúvá alakítják, majd a testre szabott XSL stíluslap felhasználásával XML fájlra. A formátum a teljes dokumentumot tartalmazó *DocBook* fájl alapján áll össze. Ha a dokumentumot XML formátumban publikálták, a szerző ismétellen ellenőrizni tudja a szöveget.

A szöveges fájlok és a metaadatok a *DiVA-portálon* keresztül szabadon elérhetők, a metaadatok újra fel is használhatók. A DiVA támogatja az *Open Archives Protocolt* és az RSS-t. Van néhány előre definiált keresőprofil, mint például a beérkező tézisek vagy a kurrens publikációk listája. A tanszékek vagy magánszemélyek saját honlapjukra kérhetik a találatokat.

A DiVA-t használó valamennyi egyetemnek saját helyi rendszere van, interfésze az egyetem grafikus felületéhez illesztett. A helyi szerver gyakran tartalmaz olyan metaadatokat, amelyeket a teljes szövegben nem publikálnak.

Tárolásra, indexelésre és keresésre kezdetben Oracle adatbázis-kezelőt használtak, ezzel párhuzamosan a metaadatokat a fájlrendszerben XML fájlokban tárolták. Helyette később egy *Apache Lucene* keresőmotor-használót helyeztek üzembe.

A DiVA dokumentumformátuma

A DiVA publikációihoz saját dokumentumformátumot fejlesztettek (*DiVA Document format = DDF*). Ez valójában egy hierarchikus, moduláris és kiterjeszhető XML sablon. A metaadatokat a strukturális adatokkal kombinálja, és lehetővé teszi a metaadatok tárolását.

A formátummal szemben két fontos követelményt támasztottak: legyen kompatibilis a többi metaadat-formátummal, és legyen emberi szemmel olvasható. A DDF jelenlegi verziója több mint száz elemet tartalmaz. Néhány elembe más elemeket ágyaztak be. A legtöbb beágyazott elem besorolási azonosítót tartalmaz, vagy például egy egyetem „telefonkönyve” a szervezeti egységek és az alkalmazottak adataival.

A hosszú távú elérhetőség

A digitális dokumentumok hosszú távú hozzáférésehez és felhasználásához folyamatosan működő azonosítóra van szükség. Erre a célra az URN:NBN-t választották. Az URN:NBN nemcsak a dokumentum azonosítójaként működik, hanem fájlnevként is: a metaadatok, a szövegfájlok és a csatolt fájlok nevében is szerepel.

Együttműködés

A DiVA konzorciális keretekben működik. Az együttműködés anyagi és egyéb feltételeit az egyes egyetemek és az Uppsalai Egyetem közötti szerződések rögzítik. A konzorciumhoz belépési díjjal lehet csatlakozni, ami a projekt fejlesztési költségvetésébe kerül. A résztvevők választhatnak, hogy helyi DiVA rendszert használnak, vagy az uppsalaihoz csatlakoznak.

A DiVA használatát a következők segítik:

- helpdesk, amelytől a rendszeradminisztrátorok folyamatos segítséget kaphatnak;
- Wiki, ahol a rendszerdokumentáció és a *gyakran ismételt kérdések (FAQ)* található; ide mindenki megírhatja észrevételeit, vagy amit mások számára is érdekesnek ítélt;
- a rendszeradminisztrátorok megkapják a rendszer használatához szükséges bevezető ismereteket;

- évente kétszer felhasználói munkaértekezletet tartanak, amelyeken különböző munkacsoportokat alakítanak, például a rendszer fejlesztésére;
- levelezőlista: nyílt csatorna az elektronikus publikálással kapcsolatos kérdések megvitatására;
- hírlevél: a legújabb fejlesztésekről, a soron lévő konferenciáról, és a DiVA-val kapcsolatos kérdésekről tájékoztat;
- konzultációs lehetőség a fejlesztésben résztvevőknek és a bekapcsolódni szándékozókknak.

A konzorciumi résztvevők mellett más intézményekkel is együttműködnek. Ennek köszönhető például a Svéd Nemzeti Könyvtárral kötött megállapodás, amelynek keretében a DiVA-ban tárolt dokumentumokat archiválták, és megszületett az URN:NBN feloldó szolgáltatás.

A kutatómunkában együttműködés alakult ki az elektronikus publikálás, az elektronikus dokumentumok hosszú távú hozzáférhetővé tétele, a tartalom megjelenítése területén.

További fejlesztések és a projekt jövője

A DiVA üzemszerű működése óta folyamatos a következő generációk tervezése. Ide sorolható az új indexelő- és keresőmodul, és a DiVA dokumentumformátumához jelenleg használt XML felváltása újabb, szabványos eszközökkel (pl. *DIDL*, *METS*),. A cél, hogy a DiVA-ban az utó- és előnyomatokat (post- and preprint) is tudják tárolni. Keresik az utat, hogy a DiVA-t „sötét archívumból” az *OAIS* modellt alkalmazó archívummá fejlesszék. További fejlesztési feladat, hogy ugyanabból az XML forrásból támogassa a rendszer az elektronikus és a hagyományos publikálást.

Nem a projekt részeként fejlesztették az elektronikus publikálás webalapú munkafolyamatát. A rendszer induláskor a dokumentumok elkészítését vezérli, ezzel írhatják a hallgatók a diplomamunkájukat, így készülhetnek a cikkek. A dokumentumokat e-mailben lehet továbbítani, például a szerző az opponenshez, vagy a folyóirat szerkesztőségéhez és vissza.

/KLOSA, Uwe – MÜLLER, Eva: The DiVA system: Current status and ongoing development. = INFO-RUM 2006: 12th Conference on Professional Information Resources, Prague, May 23-25, 2006.

http://www.inforum.cz/inforum2006/pdf/klosa_Uwe.pdf

(Wéber Katalin)