

V NTK využíváme CDS Invenio pro repozitář šedé literatury v rámci projektu NUŠL

Konfigurace CDS Invenio pro potřeby šedé literatury - NUŠL

odkaz invenio.ntkcz.cz

V NTK nyní verze 0.99.1

V horizontu měsíce (dvou) „opravdová“ verze 1.0

Virtuální prostředí VirtualBox odkaz <http://www.virtualbox.org/>.
může obsahovat jeden či více virtuálních strojů
práce z příkazové řádky i z grafické konzole

Virtuální stroj invenio.ntkcz.cz operační systém Linux-Debian

Výhody snadná přenositelnost, neboť VirtualBox umožňuje celý virtuální stroj s konfigurovaným CDS Inveniem exportovat do formátu OVF (Open Virtual Format)

potenciální příjemce už konfigurované CDS Invenio importuje do svého virtuálního prostředí VirtualBox

nezávislosti na operačním systému lokálního serveru, protože VirtualBox lze nainstalovat na všechny víceméně známé platformy

navíc se ukázalo, že exportovaný virtuální počítač lze importovat i do jiného virtuálního prostředí - VMWare ESXi.

CDS Invenio je Open Source

software pro tvorbu digitálních knihoven vyvíjený institucemi CERN (European Organization for Nuclear Research) a EPFL (École Polytechnique Fédérale de Lausanne), odkaz <http://cdsware.cern.ch/invenio/index.html>

Podpora k CDS Inveniu

zdarma, pokud jde o jednotlivé dotazy a bugy
placená, pokud CERN zajišťuje i instalaci, speciální konfiguraci atd.

Lokalizace CDS Invenio do 18 jazyků včetně češtiny ~ ne zcela

CDS Invenio využíváno různými institucemi v mnoha zemích světa

Technické info

CDS Invenio je instalovatelné na Unixových systémech

Další software: MySQL databázový server
Apache webový server
Překladač pro Python

Instalace ze zdrojového kódu ~ make & install

Protokoly a certifikáty CDS Invenio vyžaduje certifikovaný protokol https. Při publikaci CDS Invenia do vnější internetové sítě se tedy předpokládá získání veřejného certifikátu od nějaké authority

Knihovnické info

Formáty digitálních objektů libovolný formát
fulltexty stejně jako multimédia
pokud je vložen dokument v pdf formátu CDS Invenio si
jej uloží navíc jako komprimovaný postscript *.ps.gz.
jednomu záznamu může odpovídat více souborů
různých formátů nebo různých verzí

Metadata jsou vnitřně reprezentována ve formátu MARC 21, resp. MARCXML
je možné zobrazit téměř v libovolných metadatových formátech
harvestované metadatové formáty mohou být automaticky konvertovány

Sbírky libovolná struktura
možnost omezení přístupu

Prohledávání fulltextové
metadatové
podle referencí
atd.

Třídění záznamů podle různých kritérií: citace, impact faktor, atd.

Možnosti uživatelského účtu: vlastní sbírky (košíky)

Další: upozornění na nové dokumenty, diskuse, komentování dokumentů, sledování statistik, atd.

Další vlastnosti

CDS Invenio umožňuje spravovat velké repozitáře digitálních objektů
umožňuje velkokapacitní OAI harvestování
zachovává si rychlost navzdory komplexnosti služeb
poskytuje variabilitu nastavení díky své modulární architektuře

Každý modul má speciálně definovanou funkcionalitu
interaguje s ostatními moduly a databázemi

Vzájemná interoperabilita je opět konfigurovatelná dle požadavků uživatele

Modulární architektura

CDS Invenio se skládá z několika modulů, přičemž každý tento modul je vybaven specifickou funkcionalitou. S moduly lze pracovat obecně dvěma způsoby: Buď zápisem do příkazové řádky, a nebo užitím interaktivního webového rozhraní.

Moduly pro získávání a zpracování záznamů

BibConvert konverze metadat do vnitřního formátu CDS Invenia, tj. MARCXML

BibUpload upload MARCXML soubory do databáze CDS Invenia

WebSubmit definice všeho co souvisí s vkládáním metadat a dokumentů

např. vytváří formuláře a šablony pro vkládání záznamů odpovídajících různým sbírkám, dokumenty mohou podléhat také schvalování po vytvoření odpovídajícího formuláře pro danou sbírku.

BibHarvest harvestuje metadata z jiných OAI repozitářů

parametry: jméno repozitáře, periodičita harvestování, způsob konverze získaných záznamů, uploadování konvertovaných záznamů do databáze CDS Invenia

BibEdit editace už nahraných metadat

Moduly pro další zpracování a kontrolu dat

K dalšímu “jemnějšímu” zpracování dat jsou k dispozici např. tyto moduly:

BibIndex indexování metadat, tj. konfigurace tzv. „word files“, tedy která metadatová pole jsou indexována ke kterým slovním indexům. Tyto slovní indexy jsou pak užity při vyhledávání. Např. definujeme, že index „autor“ je přiřazen MARCovému poli 720\$a

BibRank třídění metadat, tj. konfigurace různých třídících metod, které mohou být užity při vyhledávání: Impact Factor, ..

BibFormat formátování metadat pro výstup: MARC, MARCXML, DC, NUŠL, BibTeX
..

WebSearch konfigurace rozhraní pro vyhledávání ve sbírkách: definice sbírek a podsbírek, definice indexů pro vyhledávání, atd.

BibClassify automatické vyhledávání klíčových slov v dokumentech pro zvolené řízené slovníky a hesláře

WebStat konfigurace statistik ukazujících využívanost jednotlivých záznamů

Administrátorské moduly

Moduly zajišťující chod a vzhled CDS Invenia

BibSched plánovač periodických úkolů vykonávaných dalšími moduly a přidělování priorit odpovídajícím procesům. Např., BibIndex, BibFormat, WebColl (update cache pro sbírky), oaiharvest (harvestování definovaných archivů), oaiarchive (Vystavení archivu invenio.ntkcz.cz pro OAI))

WebAccess definice přístupových práv k jednotlivým administrátorským či knihovnickým činnostem pro dílčí správce. Např. správa jednotlivých modulů, vkládání dokumentů, schvalování a recenzování dokumentů atd.

WebStyle definice stylu webového rozhraní CDS Invenia

Moduly pro konfiguraci uživatelských kont

Moduly umožňující personalizaci rozhraní CDS Invenia různým požadavkům koncových uživatelů jsou např. tyto:

WebSession identifikace uživatele a jeho osobní konfigurace

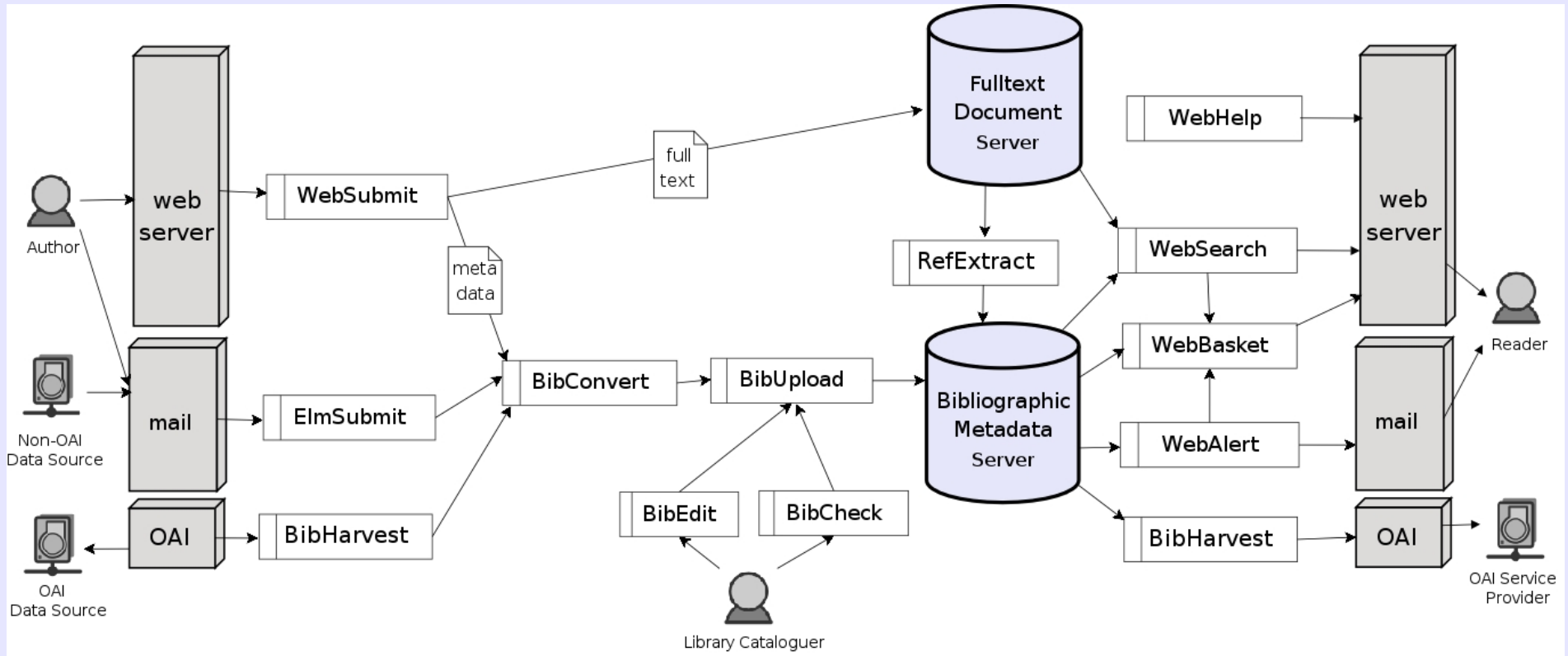
WebBasket kontrola a manipulace s množinami uživatelských košíků

WebAlert jednotlivý uživatel si definuje upozornění zasílaná emailem

WebComment manipulace s komentáři a recenzemi přidanými uživateli

„workflow“ diagram

Diagram znázorňuje spolupráci jednotlivých modulů



Děkuji

jindrich.dolansky@techlib.cz