

POLOPROVOZ ZNALOSTNÍ DATABÁZE INTERPI – DOKUMENTACE

INTERPI – Interoperabilita v paměťových institucích

Program aplikovaného výzkumu a vývoje národní
kulturní identity (NAKI)
(DF11P010VV023)

Zpracovali:

Marie Balíková (Národní knihovna)

Miroslav Kunt (Národní archiv)

Jana Šubová (Cosmotron Bohemia, s. r. o.)

Nadežda Andrejčíková (Cosmotron Bohemia, s. r. o.)



INTERPI

13. 11. 2015

Upozornění

Tato dokumentace obsahuje popis řešení poloprovozu znalostní databáze INTERPI. Část poloprovozu spočívá v technologických komponentech, jako jsou datové úložiště, aplikační a webový server, webové rozhraní apod. Další část je založena na datech získaných převodem dat ze Souborů národních autorit a dat pořízených Národním archivem při řešení projektu. Data jsou neustále doplňována a kontrolována, tak aby byla zvyšována jejich kvalita pro prezentaci a využívání databáze. Některé součásti řešení (technologie pro výměnu dat s externími systémy) vyžadují odpovídající technologické řešení na straně externího systému.

Obsah

0	Úvod	3
1	Skupiny uživatelů v poloprovozu INTERPI	4
1.1	Zpracovatelé, supervizoři a administrátoři	5
1.2	Instituce	5
1.3	Registrovaní běžní uživatelé	6
2	Technologické komponenty poloprovozu znalostní databáze INTERPI	7
2.1	Datové úložiště a databázový systém, aplikační server a webový server	7
2.2	Webové uživatelské rozhraní pro zpracování entit	8
2.3	Webové uživatelské rozhraní pro prezentaci informací o entitách	9
2.4	Webové administrátorské rozhraní pro správu uživatelských přístupů	10
2.5	Technologie pro komunikaci s jinými systémy - webové služby, Z39.50 server (s funkcí „database update“)	11
2.5.1	Proprietární XML	12
2.5.2	Z39.50	12
2.5.3	Webové služby	13
2.5.4	Využití dotazovacího jazyku PQF	13

0 Úvod

Pro zabezpečení poloprovozu znalostní databáze INTERPI bylo potřeba:

- definovat skupiny uživatelů, kteří k znalostní databázi přistupují,
- vytvořit technologickou infrastrukturu pro vytváření a prezentaci znalostní databáze,
- vytvořit metodické materiály pro vytváření znalostní databáze.

Prvé dvě části řešení jsou popsány v následujících kapitolách tohoto dokumentu. Metodické materiály jsou zpřístupněny formou certifikovaných metodik.

1 Skupiny uživatelů v poloprovozu INTERPI

Pro poloprovoz znalostní databáze INTERPI jsou definovány následující úrovně uživatelů:

- registrovaní uživatelé - pro přístup k službám znalostní databáze mají vytvořenou platnou registraci (uživatelský účet) včetně uživatelského jména a hesla:
 - zpracovatelé - vkládají a upravují záznamy prostřednictvím webového rozhraní pro zpracování entit, jsou zařazeni do skupiny podléhající konkrétnímu supervizorovi, skupiny mohou být vytvořeny podle oboru, nebo podle příslušnosti k instituci; pro tuto skupinu uživatelů jsou vytvořeny individuální účty;
 - supervizoři - kontrolují a schvalují návrhy nových záznamů entit a úpravy stávajících; pro tuto skupinu jsou vytvořeny individuální účty;
 - instituce - jde o instituce přispívající do znalostní databáze, nebo využívající informace z ní prostřednictvím některé z technologických platforem (webové služby nebo Z39.50 server); pro tuto skupinu uživatelů jsou vytvořeny skupinové účty;
 - administrátoři - úzce vymezená skupina uživatelů, která především spravuje uživatelské účty, přiděluje přístupová práva a zařazuje uživatele do skupin podléhajících konkrétním supervizorům; pro tuto skupinu uživatelů jsou vytvořeny individuální účty;
 - běžní uživatelé - odborná a laická veřejnost, která využívá rozhraní pro zpřístupnění entit a pro poskytování vybraných služeb rozhraní je nutná jejich registrace; pro tuto skupinu jsou vytvořeny individuální účty;
- anonymní uživatelé - uživatelé bez registrace, kteří mají přístup pouze k základním funkcím rozhraní pro zpřístupnění a prezentaci znalostní databáze.

Tab. 1: Tabulka uživatelských skupin registrovaných uživatelů

	primární pracovní prostředí	vkládání nových záznamů	úprava existujících záznamů	schvalování úprav záznamů	správa uživatelských účtů
individuální uživatelské účty					
administrátoři	webové rozhraní pro administraci	<i>ne</i>	<i>ne</i>	<i>ne</i>	<i>ano</i> na všech úrovních
supervizoři	webové rozhraní pro zpracování	<i>ano</i> v rámci své skupiny	<i>ano</i> v rámci své skupiny	<i>ano</i> v rámci své skupiny	<i>ne</i> (pouze své identifikační údaje)
zpracovatelé	webové rozhraní pro zpracování	<i>ano</i> v rámci své skupiny	<i>ano</i> v rámci své skupiny	<i>ne</i>	<i>ne</i> (pouze své identifikační údaje)
běžní uživatelé	webové rozhraní pro prezentaci	<i>ne</i>			
skupinové uživatelské účty					
instituce	webové služby, služby Z39.50 serveru	<i>ano</i> v rámci své skupiny	<i>ano</i> v rámci své skupiny	<i>ne</i>	<i>ne</i> (pouze své identifikační údaje)

Pro správu uživatelských účtů a přístupových práv je vytvořena centrální databáze uživatelů. Podle zařazení do skupiny uživatelů jsou vyžadována jiná data o uživateli. Individuální uživatelské účty mají vždy zabezpečen přístup k webovému rozhraní pro zpřístupnění a prezentaci bez ohledu na to, do které skupiny jsou uživatelé zařazeni.

Správa uživatelských účtů je přístupná pouze administrátorům prostřednictvím webového rozhraní pro správu uživatelských přístupů. Databáze uživatelů obsahuje pouze informace, které jsou nevyhnutelné pro dostatečnou identifikaci uživatele. Vzhledem k tomu, že je potřebné minimalizovat náročnost a pracnost správy uživatelských účtů je nutné uvést v registraci povinně e-mailový kontakt. Prostřednictvím e-mailu bude možné vygenerovat přístupové heslo a to v případě zakládání nového uživatele i v případě ztráty hesla. Přihlašovací heslo je údaj, který je v databázi uživatelů ukládán v šifrované podobě bez možnosti zpětného dešifrování a to ani ze strany administrátora. Správa uživatelského hesla je tak plně v rukách konkrétního uživatele.

Pro zvýšení bezpečnosti provozu databáze je vyžadováno vytvoření bezpečného hesla. Za bezpečné je možné považovat heslo s minimálně jednou číslicí, změnou alespoň jednoho písmena z velkého na malé, nebo obráceně a jedním speciálním znakem. Pro běžné uživatele je doporučeno heslo o délce minimálně 8 znaků, pro zpracovatele, supervizory a instituci je vyžadováno heslo minimálně v délce 10 znaků, pro administrátora v délce minimálně 15 znaků.

V případě zpracovatelů, institucí, supervizorů a administrátorů bude uživatelům po 150 dnech doporučována změna hesla.

1.1 Zpracovatelé, supervizoři a administrátoři

Pro registraci uživatele v těchto skupinách jsou potřebné tyto informace:

- jméno a příjmení,
- přihlašovací jméno a heslo,
- e-mailová adresa - jedna primární a povinná, a podle potřeby maximálně 2 další adresy,
- zaměstnavatel (nepovinný údaj) - lze předpokládat, že do řad zpracovatelů mohou být zařazeni jak pracovníci jednotlivých paměťových institucí, tak odborníci na „volné noze“.

Pro registraci uživatele ve skupině zpracovatelů je povinný i údaj:

- přiřazení supervizora - povinný údaj, který určuje oblast entit, ke kterým má zpracovatel přístup v rozhraní pro zpracování, ostatní entity může uživatel v rozhraní pro zpracování pouze prohlížet,
- omezení skupin entit (nepovinný údaj) - z oblasti entit spadající pod přiřazeného supervizora je možné konkrétnímu zpracovateli přiřadit pouze specifickou skupinu (třidu, podtřidu) entit.

Pro registraci uživatele ve skupině supervizorů je povinný údaj:

- přiřazení skupin entit - vyjmenování tříd, nebo podtříd, které supervizor spravuje v rozhraní pro zpřístupnění, ostatní skupiny entit může pouze prohlížet.

1.2 Instituce

Pro registraci uživatele ze skupiny institucí jsou potřebné tyto informace:

- název instituce,

- jméno a příjmení kontaktní osoby
- přihlašovací jméno a heslo,
- e-mailová adresa - jedna primární a povinná, a podle potřeby maximálně 2 další adresy.

1.3 Registrovaní běžní uživatelé

Pro registraci běžného uživatele jsou potřebné tyto informace:

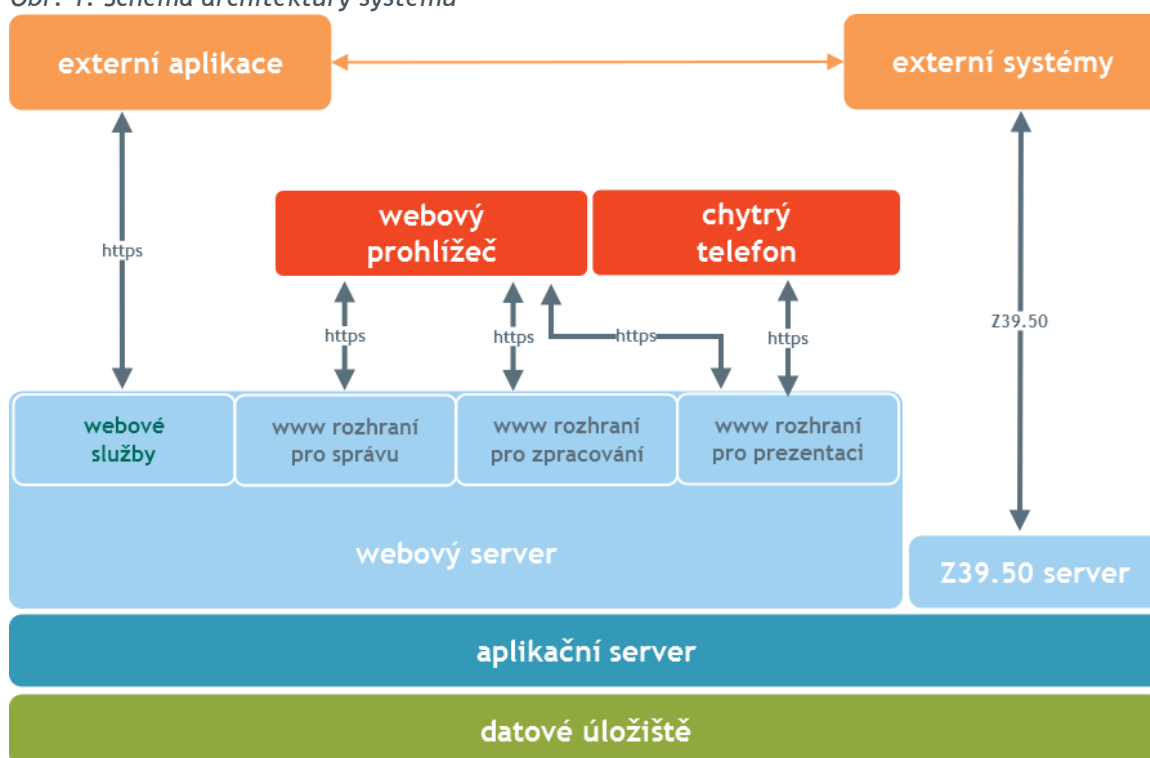
- jméno a příjmení,
- přihlašovací jméno a heslo,
- e-mailová adresa - jedna primární a povinná, a podle potřeby maximálně 2 další adresy,
- zaměstnavatel (nepovinný údaj).

2 Technologické komponenty poloprovozu znalostní databáze INTERPI

Technologickými komponenty provozu znalostní databáze INTERPI jsou:

- datové úložiště a databázový systém, aplikační server a webový server,
- webové uživatelské rozhraní pro zpracování entit,
- webové uživatelské rozhraní pro prezentaci informací o entitách,
- webové administrátorské rozhraní pro správu uživatelských přístupů,
- technologie pro komunikaci s jinými systémy - webové služby, Z39.50 server.

Obr. 1: Schéma architektury systému



2.1 Datové úložiště a databázový systém, aplikační server a webový server

Datové úložiště - volba datového úložiště byla ovlivněna více faktory:

- vícerozměrnost dat a jejich různé reprezentace,
- výkonnost,
- vývojové prostředky, platforma, interoperabilita,
- návaznost na další rozvoj,
- hardwarové nároky.

Datové úložiště je realizováno technologií InterSystems Caché®, jde o databázovou technologii nové generace, která poskytuje vysoký výkon v rychle se vyvíjejícím prostředí. Caché poskytuje

vícenásobný přístup k datům. Data jsou popsána pouze jednou v integrovaném datovém slovníku a jsou okamžitě přístupná s využitím objektů i SQL. Pro řešení jsme využili tzv. Database files (neboli „globals“), což jsou vícerozměrná pole. Tento typ úložiště nevyžaduje deklaraci a definici entit, tyto jednoduše začnou existovat tak, jak jsou využívány, a promění se na potřebné údaje o tom, co je ukládáno a využíváno. Pole není nutné specifikovat z hlediska velikosti, dimenze, typu dat atd.

Z hlediska přístupu k datům podporuje Caché tradiční indexy stejně jako bit-mapové nebo tzv. bit-slice indexy, které se využívají pro práci s daty v reálném čase. Z možností nabízených Caché jsme při realizaci využili „Transactional Bit-map Indexing“ - tj. procesní bit-mapové indexování, které zohledňuje víceúrovňovou datovou strukturu. Vytváření indexů je rychlejší a využití sofistikované metody pro komprese výrazně snižují nároky na uložení.

Pro zabezpečení vyhledávání ve webových rozhraních je implementována také technologie SOLR.

2.2 Webové uživatelské rozhraní pro zpracování entit

Rozhraní pro zpracování entit je řešeno jako webové rozhraní technologiemi ExtJS. Komunikace probíhá výlučně prostřednictvím zabezpečeného protokolu https a na přístup do rozhraní je potřeba využívat přihlašovací jméno a heslo.

Pro výměnu dat mezi rozhraním a databází jsou využívány proprietární webové služby.

Hlavní části rozhraní tvoří:

- vyhledávání entit,
- editační formulář,
- seznam otevřených záznamů.

Vyhledávání entit je řešeno vyhledávacím oknem, které umožňuje vyhledat již zpracované záznamy pomocí základních vyhledávacích kritérií (např. označení entity, označení podle jednotlivých tříd entit, hlavní část označení atd.). Vyhledávací kritéria jsou doplněna nabídkou kritérií pro zpřesnění vyhledávání, především pro vyhledání rozpracovaných záznamů, nebo záznamů čekajících na supervizi. Vyhledávací okno se využívá v úvodu zpracování, především pro ověření, zda konkrétní entita již v databázi neexistuje. Druhé využití vyhledávacího okna je v případě, kdy je potřebné vyhledat související entitu a vytvořit propojení.

Editační formuláře jsou vytvořeny pro všechny třídy entit, pro třídu entit dílo/výtvor jsou vytvořeny formuláře tři, protože je možné vymezit specifické vlastnosti především pro označení jistých skupin děl/výtvořů. Editační formuláře jsou rozděleny na skupiny údajů, logicky na sebe navazující, nebo spolu související.

třída entit / formulář	skupiny vlastností v editačním formuláři		
osoba/bytost	<ul style="list-style-type: none"> formy označení jiná identita stručná charakteristika životopisné údaje zařazení 	<ul style="list-style-type: none"> rodinné/partnerské vztahy aktivity ocenění další vztahy poznámky 	<ul style="list-style-type: none"> obrázky prameny, autoři jiné ID administrativní údaje



korporace	<ul style="list-style-type: none"> formy označení historie sídlo popis a funkce normy 	<ul style="list-style-type: none"> struktura členství vlastnické vztahy zařazení další vztahy 	<ul style="list-style-type: none"> poznámky obrázky prameny, autoři jiné ID administrativní údaje
geografický objekt	<ul style="list-style-type: none"> formy označení popis historie zařazení souřadnice/kódované údaje 	<ul style="list-style-type: none"> hierarchická struktura vlastnické vztahy další vztahy poznámky 	<ul style="list-style-type: none"> obrázky prameny, autoři jiné ID administrativní údaje
rod/rodina	<ul style="list-style-type: none"> formy označení historie a jiné stručná charakteristika zařazení 	<ul style="list-style-type: none"> rodinné vztahy další vztahy poznámky obrázky 	<ul style="list-style-type: none"> prameny, autoři jiné ID administrativní údaje
událost	<ul style="list-style-type: none"> formy označení místo a organizace historie stručná charakteristika 	<ul style="list-style-type: none"> zařazení další vztahy poznámky obrázky 	<ul style="list-style-type: none"> prameny, autoři jiné ID administrativní údaje
autorská a umělecká díla	<ul style="list-style-type: none"> formy označení popis 	<ul style="list-style-type: none"> vlastnické vztahy zařazení 	<ul style="list-style-type: none"> obrázky prameny, autoři
stavby	<ul style="list-style-type: none"> historie 	<ul style="list-style-type: none"> další vztahy 	<ul style="list-style-type: none"> jiné ID
ostatní díla	<ul style="list-style-type: none"> umístění, lokalita 	<ul style="list-style-type: none"> poznámky 	<ul style="list-style-type: none"> administrativní údaje
obecný pojem	<ul style="list-style-type: none"> formy označení vymezení pojmu pokyny k použití historie 	<ul style="list-style-type: none"> zařazení kódované údaje další vztahy poznámky 	<ul style="list-style-type: none"> obrázky prameny, autoři jiné ID administrativní údaje

Vybrané záznamy jsou otevřeny na záložkách webového rozhraní, je tak možné editovat více záznamů najednou. Pro lepší orientaci slouží seznam otevřených záznamů, ve kterém je možné záznamy hromadně zavřít.

Uživatelská příručka pro práci s rozhraním je k dispozici jako součást webového rozhraní a taktéž na stránce www.interpi.cz v sekci s výsledky projektu.

2.3 Webové uživatelské rozhraní pro prezentaci informací o entitách

Prostředí pro vyhledávání a prezentaci informací o entitách poskytuje přístup dvěma způsoby: vyhledáváním pomocí formuláře pro jednoduché a rozšířené vyhledávání s výběrem různých selekčních kritérií, nebo procházením jednotlivých tříd a jejich podtříd.

Webové rozhraní pro zpřístupnění a prezentaci znalostní databáze je realizováno pomocí technologií CSP (Caché Server Pages), které jsou integrální součástí databázového systému.

Stránky jsou tvořeny v značkovacím jazyce HTML, pro úpravu vzhledu stránek jsou aplikovány CSS (kaskádových stylů), s kterými souvisí další technologie, již je LESS - Stylesheet language neboli CSS preprocessor, kterým ze zdrojového kódu zapsaného ve vlastní syntaxi generujeme CSS pro prohlížeč. LESS zavádí do CSS některé vlastnosti, které známe např. z programovacích jazyků, jako jsou proměnné, výrazy nebo makra. Z hlediska přístupnosti vytvořených stránek pro uživatele se specifickými potřebami je aplikována technologie WAI-ARIA - Web Accessibility Initiative - Accessible Rich Internet Applications).

Protože je zabezpečen přímý přístup rozhraní do databáze, data jsou zpřístupňována bez prodlení.

Uživatelská příručka pro práci s rozhraním bude po schválení průběžné správy o řešení projektu za rok 2015 k dispozici na stránce www.interpi.cz v sekci s výsledky projektu.

2.4 Webové administrátorské rozhraní pro správu uživatelských přístupů

Rozhraní pro správu uživatelských přístupů je určeno úzké skupině uživatelů - technologicky je řešeno stejným způsobem jako rozhraní pro zpracování dat.

Služby rozhraní:

- vyhledání uživatele,
- změna údajů o uživateli (jméno, příjmení, e-mail, příp. zaměstnavatel),
- nastavení skupiny uživatelů, přiřazení supervizora, omezení skupiny entit,
- odeslání kontaktního e-mailu pro vygenerování hesla.

Rozhraní je v době odevzdání průběžné zprávy za rok 2015 v průběhu realizace.

V současné době není v projektu INTERMI řešena možnost sdílení uživatelů a jejich práv z jiných institucí a systémů; na úrovni rezortu zatím nedošlo ke sjednocení v přístupu uživatelů tak, aby byl zabezpečen přístup k různým aplikacím, které jsou pro ně určené.

2.5 Technologie pro komunikaci s jinými systémy - webové služby, Z39.50 server (s funkcí „database update“)

Obr. 2: Schéma použití technologií pro komunikaci s jinými systémy



Důležité je zmínit formát prezentace dat, který je využíván pro komunikaci s externími systémy. V případě protokolu Z39.50 je formátem pro získání dat MARC (primárně MARC21 - formát pro authority). Zde je nezbytné počítat s určitou mírou ztráty dat, neboť struktura informací o entitě převyšuje strukturu formátu MARC. Pro prezentaci prostřednictvím webových služeb může být využita jak interpretace MARC do XML (tj. MARC/XML) - se stejnou mírou ztráty dat, tak především proprietární XML využité pro projekt INTEPRI.

XML - eXtensible Markup Language - neboli rozšiřitelný značkovací jazyk lze považovat za formát dokumentů, který používá značky pro vyjádření obsahu. Značky nejsou předem definovány a lze je libovolně přidávat.

Realizace INTERPI počítá s prezentací dat v různých formátech i v případě, že je lze využít jenom pro vymezenou skupinu entit (jednotlivé třídy nebo více tříd), a to především pokud se tím podpoří prezentace v prostředí sémantického webu.

Aplikace RDF vhodná pro tezaury a klasifikační systémy je SKOS - Simple Knowledge Organization System. Umožňuje vyjádřit strukturu a obsah konceptuálních schémat - tezurů, klasifikačních schémat, seznamů předmětových záhlaví, taxonomií, folksonomií a jiných typů kontrolovaných slovníků. SKOS je možné využít jako alternativu pro prezentaci entit především

ze třídy obecný pojem. Jiné využití RDF je prezentováno ve FOAF - ontologii pro popis osob, jejich aktivit a vztahů.

2.5.1 Proprietární XML

Pro potřeby zpřístupnění dat pro externí systémy jsou vytvořeny dvě podoby proprietárního XML:

- kompletní informace,
- stručné informace.

Proprietární XML vychází z vymezení obecného konceptuálního modelu, pro pojmenování prvků byla zvolena čeština a cílem bylo poskytnout informace v co nejvyšší míře v přirozeném jazyce.

Kompletní verze proprietárního XML obsahuje všechny informace o entitách, ze všech tříd.

Stručná verze proprietárního XML obsahuje pouze vybrané informace o entitách a slouží především pro základní identifikaci entity.

Pro definici XML je využito XSD publikováno na stránce www.interpi.cz.

2.5.2 Z39.50

Z hlediska interoperability sehrávají významnou roli v realizaci INTERPI nástroje pro komunikaci. V prostředí knihoven jím je především již déle užívaný protokol Z39.50 neboli norma ISO23950. Jde o protokol, na principu komunikace klient - server, který definuje průběh a způsob komunikace při vyhledávání v databázi a při získání výsledů (odpovědi).

Komunikaci zahajuje vždy klient. Pomocí IP adresy, portu, názvu databáze, popřípadě uživatelského jména a hesla iniciuje spojení se serverem. Po jeho úspěšném navázání může proběhnout samotné vyhledávání a získání výsledků. Při vyhledávání se využívá množina atributů známá pod označením BIB1 atribut set vymezující zjednodušeně, jakým způsobem se klient dotazuje (jaké selekční kritérium je použito, jaký typ rozšíření je použit apod.). Množinu atributů typu „use“ jsme rozšířili o specifická selekční kritéria, která reflektují rozsah a strukturu dat v záznamu INTERPI. Dotaz je odevzdáván v jazyku PQF (Prefix Query Format), který se prakticky stal standardem pro aplikaci standardu Z39.50.

V projektu INTERPI byla ze standardu Z39.50 aplikována nejenom část pro vyhledávání informací, ale také část pro vkládání a modifikaci dat, protože to vyžaduje současná praxe v oblasti knihoven.

Údaje o Z39.50 serveru

adresa: 195.113.132.114

databáze: interpi_us_auth

formát: USMARC (tj. MARC21)

Přihlašovací údaje získává po registraci instituce.

Poznámka: aktuální informace o připojení k Z39.50 serveru jsou na stránce www.interpi.cz.

2.5.3 Webové služby

Ostatní paměťové instituce upřednostňují komunikaci pomocí webových služeb, které implementují do využívaných systémů. Webové služby představují rozhraní pro servisně orientovanou architekturu (SOA), ve které (především) webové aplikace mohou dynamicky spolupracovat s jinými s využitím otevřených standardů zahrnujících XML nad HTTP, UDDI a SOAP. Webové služby tudíž umožňují komunikaci na bázi protokolu HTTP, který je na rozdíl od Z39.50 bezstavový - server po odeslání odpovědi ukončí spojení s klientem. Existují dva způsoby pro výměnu správ mezi klientem a serverem - SOAP (Simple Object Access Protocol) nebo REST (Representational State Transfer).

Prostřednictvím webových služeb mohou jiné systémy nejenom vyhledat informace o entitách, ale také vytvářet a modifikovat záznamy ve znalostní databázi INTERPI. Rozhraní jednotlivých webových služeb je popsáno prostřednictvím jazyka WSDL - Web Service Description Language. Ten je založen na jazyku XML a popisuje syntaxi volání webových služeb. Tím se klientský program může připojit k dané webové službě a vyčíst informace o tom, jaké funkce daná služba poskytuje a jak je provolat. Pro formulaci dotazů prostřednictvím webových služeb se také využívá PQF.

Dokumentace webových služeb je zveřejněna na stránce www.interpi.cz.

2.5.4 Využití dotazovacího jazyku PQF

Struktura PQF dotazu

Dotazovací jazyk PQF je pro potřeby INTERPI implementován v následujícím rozsahu:

- atribut 1 (use - vyhledávací kritérium),
- atribut 4 (fráze),
- atribut 5 (rozšíření),
- atribut 6 (úplnost - completeness).

Vyhledávací termín je vždy uzavřen v uvozovkách a nejsou rozlišována malá/velká písmena.

Atribut 1 (vyhledávací kritérium)

Pro INTERPI jsou využívány převážně proprietární use atributy, které zohledňují specifika datové struktury. Atributy mohou být doplňovány nebo jejich obsah může být mírně modifikován na základě potřeb uživatelů.

Přehled implementovaných vyhledávacích kritérií:

- 2050 - stav záznamu - obsahuje kód pro stav záznamu vzhledem na jeho životný cyklus:
 - nr - rozpracovaný návrh

- nd - dokončený návrh
- rr - rozpracovaná revize
- rd - dokončená revize
- sz - schválený záznam
- zz - zrušený záznam
- ir - bez revize INTERPI (záznamy získané importem z jiných systémů)
- 2051 - třída - obsahuje jednoznakový kód pro označení třídy entit:
 - o - třída osoba/bytost
 - r - třída rod/rodina
 - k - třída korporace
 - g - třída geografický objekt
 - u - třída událost
 - d - třída dílo/výtvar
 - p - třída obecný pojem
- 2054 - preferované označení entity - obsahuje všechny preferované označení entity bez ohledu na aplikovaná pravidla a bez ohledu na zařazení do třídy - při použití jako kompletního termínu obsahuje hlavní, vedlejší část označení a doplňky
- 2055 - všechny označení entity - obsahuje všechny označení entity bez ohledu na aplikovaná pravidla, jazyk nebo preferované/variantní označení a zařazení do třídy - použití jako kompletního termínu viz atribut 2054
- 2056 - preferované označení entit třídy osoba/bytost - obsahuje všechny preferované označení entity bez ohledu na aplikovaná pravidla označení - použití jako kompletního termínu viz atribut 2054
- 2057 - všechny označení entity třídy osoba/bytost - obsahuje všechny označení entity bez ohledu na aplikovaná pravidla, jazyk nebo preferované/variantní označení - použití jako kompletního termínu viz atribut 2054
- 2058 - preferované označení entit třídy rod/rodina - obsahuje všechny preferované označení entity bez ohledu na aplikovaná pravidla označení - použití jako kompletního termínu viz atribut 2054
- 2059 - všechny označení entity třídy rod/rodina - obsahuje všechny označení entity bez ohledu na aplikovaná pravidla, jazyk nebo preferované/variantní označení - použití jako kompletního termínu viz atribut 2054
- 2060 - preferované označení entit třídy korporace - obsahuje všechny preferované označení entity bez ohledu na aplikovaná pravidla označení - použití jako kompletního termínu viz atribut 2054
- 2061 - všechny označení entity třídy korporace - obsahuje všechny označení entity bez ohledu na aplikovaná pravidla, jazyk nebo preferované/variantní označení - použití jako kompletního termínu viz atribut 2054
- 2062 - preferované označení entit třídy událost - obsahuje všechny preferované označení entity bez ohledu na aplikovaná pravidla označení - použití jako kompletního termínu viz atribut 2054
- 2063 - všechny označení entity třídy událost - obsahuje všechny označení entity bez ohledu na aplikovaná pravidla, jazyk nebo preferované/variantní označení - použití jako kompletního termínu viz atribut 2054

- 2064 - preferované označení entit třídy geografický objekt - obsahuje všechny preferované označení entity bez ohledu na aplikovaná pravidla označení - použití jako kompletního termínu viz atribut 2054
- 2065 - všechny označení entity třídy geografický objekt - obsahuje všechny označení entity bez ohledu na aplikovaná pravidla, jazyk nebo preferované/variantní označení - použití jako kompletního termínu viz atribut 2054
- 2066 - preferované označení entit třídy dílo/výtvar - obsahuje všechny preferované označení entity bez ohledu na aplikovaná pravidla označení - použití jako kompletního termínu viz atribut 2054
- 2067 - všechny označení entity třídy dílo/výtvar - obsahuje všechny označení entity bez ohledu na aplikovaná pravidla, jazyk nebo preferované/variantní označení - použití jako kompletního termínu viz atribut 2054
- 2068 - preferované označení entit třídy obecný pojem - obsahuje všechny preferované označení entity bez ohledu na aplikovaná pravidla označení - použití jako kompletního termínu viz atribut 2054
- 2069 - všechny označení entity třídy obecný pojem - obsahuje všechny označení entity bez ohledu na aplikovaná pravidla, jazyk nebo preferované/variantní označení - použití jako kompletního termínu viz atribut 2054
- 2008 - obecný doplněk - obsahuje obecný doplněk v označení entit, kde je obecný doplněk využíván
- 2009 - chronologický doplněk - obsahuje obecný doplněk v označení entit, kde je chronologický doplněk využíván
- 2010 - geografický doplněk - obsahuje obecný doplněk v označení entit, kde je geografický doplněk využíván
- 2011 - příslušnost k zemi původu - obsahuje hlavní část označení (preferovaných i variantních) entity geografický objekt, která je využita pro vyjádření příslušnosti k zemi původu
- 2012 - obor působnosti - obsahuje hlavní část označení (preferovaných i variantních) entity obecný pojem, která je využita pro vyjádření oboru působnosti
- 2013 - kategorie - obsahuje hlavní část označení (preferovaných i variantních) entity obecný pojem, která je využita pro vyjádření kategorie entity
- 2014 - jazyková oblast - obsahuje kód jazykové oblasti entity
- 2015 - datum vzniku - obsahuje datum vzniku entity
- 2016 - datum zániku - obsahuje datum zániku entity
- 2017 - místo vzniku - obsahuje hlavní část označení (preferovaných i variantních) entity geografický objekt, která je využita pro vyjádření místa vzniku entity
- 2018 - místo zániku - obsahuje hlavní část označení (preferovaných i variantních) entity geografický objekt, která je využita pro vyjádření místa zániku entity
- 2019 - sídlo - obsahuje hlavní část označení (preferovaných i variantních) entity geografický objekt, která je využita pro vyjádření sídla entity
- 2020 - geografická působnost - obsahuje hlavní část označení (preferovaných i variantních) entity geografický objekt, která je využita pro vyjádření geografické působnosti entity
- 1016 - vše - obsahuje údaje ze všech částí záznamu
- 12 - INTERPI identifikační číslo záznamu

Aktuální seznam podporovaných atributu bude zveřejněn na stránce www.interpi.cz.

Atribut 4 (fráze)

Implementovaná pouze hodnota 1 - fráze. Při použití tohoto atributu je možné zadat jako vyhledávací termín více slov - oddělených mezerami. Výsledkem budou ty záznamy, které mají ve vyhledávacím kritériu uvedené všechny zadané slova ve stejném pořadí a bez výskytu jiného slova mezi nimi. Může se před nimi a za nimi vyskytovat jakýkoliv počet jiných slov.

Atribut 5 (rozšíření)

Implementovaná pouze hodnota 1 - pravostranné rozšíření. Při použití tohoto atributu je zadaný vyhledávací termín rozšířen o jakýkoliv počet libovolných znaků. Atribut je možné využít pouze s jednoslovnými termíny - tj. není implementován pro použití s frází.

Atribut 6 (úplnost - completeness)

Implementovaná pouze hodnota 3 - kompletní pole (resp. hodnota 2 - kompletní podpole, která je používána ve stejném významu). Při použití tohoto atributu jsou výsledkem pouze záznamy, které obsahují v příslušném vyhledávacím kritériu uvedenou hodnotu přesně a úplně.

Syntaxe

Pro zadávání dotazu platí „dědičnost“ atributů. Atributy uvedené na začátku dotazu platí pro všechny termíny.

```
@attr 1=1016 @and 'xx' 'yy'
```

Atribut 1 s hodnotou "1016" platí pro oba termíny 'xx' i 'yy'.

```
@attr 6=3 @attr 1=1016 @and @and 'xx' 'yy' 'zz'
```

Atribut 6 s hodnotou "3" a atribut 1 s hodnotou "1016" platí pro všechny tři termíny 'xx' i 'yy' i 'zz'.

Pro některý z vyhledávacích termínů je možné zděděný atribut libovolně změnit, nebo doplnit jiný.

```
@attr 1=1016 @and 'xx' @attr 5=1 'yy'
```

Atribut 1 s hodnotou "1016" platí pro oba termíny, pro druhý termín platí i atribut 5 s hodnotou "1".

Lze kombinovat i více různých hodnot atributů.

```
@attr 1=1016 @and 'xx' @attr 1=2050 'yy'
```

Atribut 1 s hodnotou "1016" platí pro oba termíny, pro druhý termín platí i atribut 5 s hodnotou "1".

Pro kombinaci různých vyhledávacích termínů je možné využít operátory AND, OR, NOT.

Příklady použití

Pro následující příklad předpokládejme následující hodnoty vyhledávacího kritéria 2056 (kompletní termíny):

```
jindrák, františek 1915- (ID 001)
```

jindrák, jindřich 1931-1993	(ID 002)
jindra, jan nepomucký 1863-1930	(ID 003)
jindra, jan 1877-	(ID 004)
jindra, jan 1922-1979	(ID 005)
jindra, jan 1962-	(ID 006)
jindr, josef	(ID 007)
jindrová, zdeňka 1910-	(ID 008)
nepomucký, jan 1895-1948	(ID 009)

Vyhledávací kritérium 2056 (jednotlivá slova):

františek	(ID 001)
jan	(ID 003, 004, 005, 006, 009)
jindr	(ID 007)
jindra	(ID 003, 004, 005, 006)
jindrák	(ID 001, 002)
jindrová	(ID 008)
jindřich	(ID 002)
josef	(ID 007)
nepomucký	(ID 003, 009)
zdeňka	(ID 008)

A hodnoty kritéria 2012:

česká literatura	(ID 006, ID 008)
------------------	------------------

Příklady použití operátorů:

```
@attr 1=2056 @and 'jindra' 'jan'
výsledkem jsou ID: 003, 004, 005, 006
@attr 1=2056 @not 'jindrák' 'františek'
výsledkem je ID: 002
@attr 1=2056 @or 'jindra' 'jan'
výsledkem jsou ID: 003, 004, 005, 006, 009
```

Příklady atributu fráze:

```
@attr 1=2056 @attr 4=1 'jindra, jan'
výsledkem jsou ID: 003, 004, 005, 006
@attr 1=2056 @attr 4=1 'jan nepomucký'
výsledkem je ID: 003
```

Příklady atributu úplnost:

```
@attr 1=2056 @attr 6=3 'jindra, jan'
výsledkem není žádný záznam
@attr 1=2056 @attr 6=3 'jindra, jan 1962-'
výsledkem je ID: 006
```

Příklady atributu rozšíření:

```
@attr 1=2056 @attr 5=1 'jindr'
výsledkem jsou ID: 001, 002, 003, 004, 005, 006, 007, 008
@attr 1=2056 @attr 5=1 'jindr' 'j'
výsledkem jsou ID: 002, 003, 004, 005, 006, 007
```

Příklady použití více atributů:

```
@and @attr 1=2056 'jindra' @attr 1=2012 @attr 4=1 'česká literatura'
výsledkem je ID: 006
```