

PAPYRUS

Rastislav Černay
Ročník: tretí
Dátum spracovania: 3.12.2013

PAPYRUS

www.ict-papyrus.eu

Stručná charakteristika projektu

Projekt PAPYRUS je portál, ktorý zabezpečuje multiplatformové zbieranie informácií, ktoré ďalej ukladá a usporiada podľa dátumu vzniku a obsahu. Snaží sa byť dynamickou digitálnou knižnicou, ktorá rozumie dotazu od používateľa a dokáže vyhľadať zmysluplné výsledky aj z oblastí, ktoré pre neho nie sú domáce. Výsledky ktoré papyrus vráti sa menia podľa výsledkov z predchádzajúcich vyhľadávaní.

Doba riešenia, aktuálny stav

Cena projektu: 2 952 364 EUR

Suma financovaná EU: 2 200 000 EUR

Projekt odštartoval v roku 2008.01.03 a bol úspešne dokončený v roku 2011.21.01.

Koordinátor projektu:

Athens Technology Center SA

Účastníci:

Stichting Historie der Techniek -Holandsko

Consorzio Interuniversitario Cineca -Taliansko

Deutsche Welle -Nemecko

Agende France Presse - Francúzsko

Universita Degli Studidi Trento -Taliansko

National and Kapodistrian University of Athens - Grécko

Queen Mary and Westfield College, University of London

Ciele projektu

Hlavný cieľ projektu PAPYRUS bolo vyvinúť software potrebný pre chod dynamickej digitálnej knižnice, umožňujúcej prístup k informáciám užívateľom z najrôznejších oborov. Uľahčiť používateľovi vyhľadanie vedeckých ako aj laických publikácií.

V snahe zabezpečiť cieľ projektu sa PAPYRUS zamerá na:

- Porozumenie multimediálnym súborom za pomoci už existujúcich metód a rozpoznávania vzorových modelov.
- Vyvinutie vyhľadávača, ktorý rozumie užívateľovým vstupom.
- Implementovanie nástrojov na automatizáciu procesu mapovania vedomostí pre rovnaké pojmy medzi rôznymi zdrojmi a užívateľskými vstupmi.

- Vyvinutie techniky prezentovania a prinášanie výsledkov v zrozumiteľnej forme pre koncového užívateľa.

Popis projektu a jeho výsledky

PAPYRUS predstavuje systém viacúčelovej dynamickej digitálnej knižnice, ktorá rozumie vyhľadávaniam užívateľa, dokáže vyhľadať informácie v cudzom obore a vrátiť výsledok v užitočnom a zrozumiteľnom tvare užívateľovi. Základné kroky, ktoré sa používajú v projekte Papyrus pre vyhľadávanie v archíve sú:

1. Formulácia vstupu

Zadáme otázku alebo kľúčové slovo, o ktorom si myslíme, že je relevantné. Ak je to dôležité tak aj časový interval v ktorom mohla byť novinka vydaná.

2. Analýza a obohatenie vstupu

Naša otázka alebo kľúčové slovo sú analyzované. Ako ďalší krok, PAPYRUS použije históriu užívateľských vyhľadávaní a pomocou nej vstup obohatí o ďalšie pojmy (synonymá, rôzne príbuzné pojmy, archaizmy atď.). Tento obohacovací proces je cenný najmä v kontexte archívnych materiálov, kde sa terminológia časom menila. Napríklad slovo ‚computer‘ znamenalo niekedy ľudský robotník.

3. Dolad'ovanie vstupu

Po analýze vstupu a jeho obohatení ho PAPYRUS môže vyhodnotiť. Máme možnosť sa pozrieť na pôvodný vstup spolu s príbuznými pojmi v šablónovom tvare alebo ho reprezentovať viac doslovne. Môžeme pridať alebo odobrať pojmy, alebo nechať vstup v tvare taký aký je a tento krok úplne preskočiť.

4. Vyhľadávanie

V tejto fáze je obohatený vstup použitý na lokáciu relevantných dokumentov vo všetkých dostupných archívoch. Tieto dokumenty môžu byť textové ale môžu to byť aj obrázky, zvukové súbory alebo videá, ktoré sú indexované Papyrusom. (Indexovanie, hoci automatizované, má veľkú mieru úspešnosti, hlavne vďaka tomu, že využíva metadáta, ktoré predtým zadal človek manuálne). Týmto spôsobom sa snaží PAPYRUS dosiahnuť maximálnu presnosť pri vracaní výsledkov.

5. Prezentácia výsledkov

Po dokončení vyhľadávania máme k dispozícii viacero možností. Prvá (jednoduchšia) možnosť je nechať si zobrazit' dokumenty jeden po druhom, ako

ohodnotený list so zvýraznenými relevantnými sekciami. Druhou variáciou je zobrazenie výsledkov, ktoré sú zoradené do skupín podľa interpretácie vstupu. Užívateľ si tiež môže vybrať obor, prípadne kombináciu oborov z ktorých chce zobraziť mať zobrazené dokumenty. Výhodou je, že tieto výsledky budú použité pri ďalších podobných vyhľadávaniach, čím sa systém efektívne zdokonaľuje.

Aby toto všetko fungovalo, Papyrus priniesol nasledujúce riešenia (podčiarknuté názvy sú dostupné ako open-source):

- *The PAPHYRUS semantic multimedia analysis components* zdokumentováva obsah multimediálnych súborov. Tento program obsahuje audio analýzu, sémantickú video analýzu, textovú analýzu a mapovanie pojmov.

- *The PAPHYRUS Ontology Tools* poskytujú všetky funkcie pre porozumenie dvoch rozdielnych domén, definovanie sémantickej korešpondencie medzi nimi a sémantickú interpretáciu užívateľských vstupov podľa užívateľskej histórie.

Tieto nástroje zahŕňajú:

- The Ontology Development Component (TRENDS)* sa používa na zrozumiteľné upravovanie a mapovanie. Obsahuje mapovacie nástroje, ktoré budú slúžiť ako softwarové prototypy pre zrozumiteľnejšie mapovanie a prirad'ovacie nástroje medzi doménami histórie a noviniek.

- The Ontology Browsing Component* sa používa pre vyhľadávanie významov a predkladanie otázok v tejto oblasti. Zahŕňa analýzy otázok a výpočetný modul, ktorý transformuje otázku používateľa do viacvýznamovej formy, takže sa bude dať použiť vo viacerých repozitároch rovnako ako aj v prezentačných technikách, ktoré sú prisudzované výsledkom používateľovi v zrozumiteľnom a efektívnom tvare.

- Porozumenie obsahu Noviniek a používateľskej Históriae.

- *The PAPHYRUS Digital Library Engine* integruje všetky funkcie poskytované projektom PAPHYRUS. Vyhľadávací modul je základom, na ktorom sa portál stavia a v podstate môže byť použitý pre tento účel každý štandardný vyhľadávač. V aktuálnej implementácii je použitý open-source LUCENE SOLR vyhľadávač.

Vlastné zhodnotenie

Papyrus sa zamerá na premostenie viacerých disciplín a napomáhaniu dostupnosti ku vedomostiam, ktoré boli ľuďom z iných oborov nedostupné. Skrze rozvíjanie systému porozumenia novým veciam a zároveň aj histórie by mohol predstavovať silný nástroj pre získavanie informácií pre medziodborovú spoluprácu a rýchly prístup k pohľadu na veci z iného uhlu.

Zásadný problém papyrusu však nespočíva v technickom vypracovaní, ale v tom, že sa nepoužíva. Veľká časť tohto projektu je napísaná ako open-source a je možné, že hoci ako celok nenašiel využitie, jeho časti niekde dobre slúžia.

Jediná stránka, ktorá uvádzala, že používa POPYRUS bola www.digibic.eu, ktorá však v čase tvorby tohto projektu bola už nedostupná a bola vytvorená ako projekt EU, ktorý mal implementovať rôzne projekty týkajúce sa digitálnych knižníc dokopy. To, že už uplynuli 3 roky od ukončenia projektu iba podčiarkuje fakt, že v prípade ak niekto bude chcieť využiť túto technológiu už môže byť zastaralá, keďže projekt sa už neaktualizuje.

Zoznam literatúry

<http://www.ict-papyrus.eu>

<http://www.cordis.europa.eu/>

<http://disi.unitn.it/~velgias/docs/KatiforiNPITKSTTBKTV11.pdf>

Metadata v DC

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
```

```
<meta name="DC.Title" content="Papyrus" />
```

```
<meta name="DC.Creator" content="Rastislav Cernay" />
```

```
<meta name="DC.Subject" content="Papyurs" />
```

```
<meta name="DC.Date" content="3.12.2013" />
```

```
<meta name="DC.Type" content="Text" />
```

```
<meta name="DC.Type" content="Esej" />
```

```
<meta name="DC.Format" content="application/pdf" />
```

```
<meta name="DC.Format" content="computerFile" />
```

```
<meta name="DC.Source" content="http:// http://www.ict-papyrus.eu/" />
```

```
<meta name="DC.Source" content="http:// http://www.cordis.europa.eu/" />
```

```
<meta name="DC.Language" content="sk" />
```