

# Studentská Vědecká Konference 2011

## JMZW: VÍCEÚČELOVÉ GRAFICKÉ UŽIVATELSKÉ ROZHRAŇÍ

Dan PRESSL<sup>1</sup>

### 1 ÚVOD

V dnešní době obrovského rozvoje digitálních archivů je více než zapotřebí zajistit kvalitní zpracovávání dat, jejich prezentaci a vyhledávání v nich. Mým úkolem bylo vytvoření grafického uživatelského rozhraní pro knihovnu Voiar. Knihovna vznikla za účelem sjednocení přístupového rozhraní k různým druhům databází ve výzkumu na Katedře kybernetiky Západočeské univerzity v Plzni. V knihovně Voiar je možno ukládat textové záznamy, multimediální záznamy aj. Pro projekt JMZW, zaměřený na vytvoření systému pro automatickou aktualizaci a adaptaci slovníků a jazykových modelů pro ASR s využitím dat z internetu, bylo nutné vytvořit grafické ovládací prvky pro sadu algoritmů využívaných v JMZW. Na začátku roku 2011 se v databázi projektu JMZW nacházelo 3 085 354 článků, 3 758 615 slov a 946 993 568 tokenů[Švec, Hoidekr, Soutner, Vavruška (2011)].

### 2 POUŽITÉ TECHNOLOGIE

Celé grafické uživatelské rozhraní je postaveno na vývojářské knihovně Qt verze 4 a programovacím jazyku Python. [Guido van Rossum (2011)]Python je dynamický objektově orientovaný interpretovaný programovací jazyk. Python je dostupný pro většinu běžně používaných platforem.[Nokia (2010)]Qt je multiplatformní vývojářská knihovna, primárně určená pro vytváření grafického uživatelského rozhraní. Je podporována na většině používaných platforem.

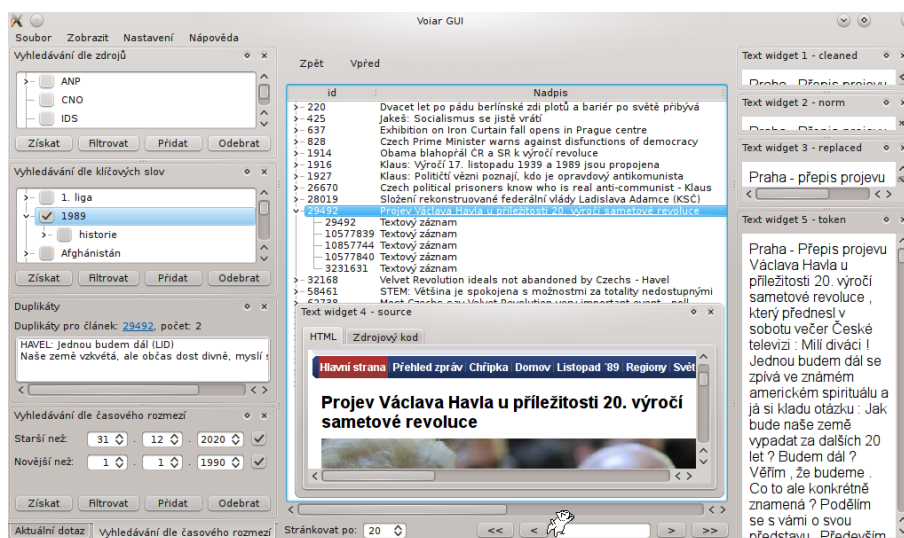
### 3 POPIS UŽIVATELSKÉHO ROZHRAŇÍ

Grafické uživatelské rozhraní se dá rozložit do dvou částí. První částí je centrální ovládací prvek (angl. central widget), který slouží k zobrazování záznamů z databáze. Pro zobrazení je použita stromová struktura, která umožňuje detailnější pohled na požadovaná data. Druhou částí, více modifikovatelnou, jsou grafická uživatelská rozhraní jednotlivých algoritmů. Ty se nachází okolo centrálního ovládacího prvku. Každý algoritmus, který vyžaduje interakci s uživatelem, má své vlastní grafické uživatelské rozhraní. Díky velkému objemu dat, která se mohou potencionálně zobrazovat, se při každém novém dotazu provádí stránkování dat, díky čemuž se vždy vypisuje požadovaný počet záznamů na stránku a odezvy rozhraní jsou velmi rychlé. Součástí grafického uživatelského rozhraní je také historie a je tedy možné si prohlížet staré výsledky vyhledávání.

#### 3.1 JMZW algoritmy

Sada algoritmů pro projekt JMZW je uzpůsobena převážně pro zobrazování textových dat. Grafické uživatelské rozhraní obsahuje algoritmy na filtrování záznamů v databázi podle zdrojů, odkud byla data získána, podle času, kdy byla data získána, podle klíčových slov, která byla daným záznamům přiřazena, podle jazyka v jakém je článek napsán, a zobrazuje nalezené duplikáty pro označený záznam v centrálním ovládacím prvku. Podle

<sup>1</sup>Bc. Dan Pressl, student navazujícího studijního programu Aplikované vědy a informatika, obor Kybernetika a řídicí technika, e-mail: pressl@students.zcu.cz



Obrázek 1: Grafické uživatelské rozhraní se sadou algoritmů pro projekt JMZW

počtu atributů daného záznamu se automaticky vygenerují i zobrazovací grafické prvky, které zobrazují data, která jsou v záznamu uložena. Během generování se automaticky rozpozná typ dat a načte se příslušný grafický prvek. Grafické ovládací prvky pro algoritmy se dají rozdělit do dvou skupin:

- **aktivní** - Každý takový prvek obsahuje 4 tlačítka, Získat, Filtrovat, Přidat a Odebrat, pro vytváření dotazu nad databází, který si mezi sebou jednotlivé grafické algoritmy sdílejí. Díky tomu je možné vytvářet komplexní dotaz, který může obsahovat všechny filtry popsané výše.
- **pasivní** - Každý takový prvek je uzpůsoben pouze k zobrazování výsledku daného algoritmu. Do této skupiny patří například algoritmus pro zobrazování nalezených duplikátů.

## 4 ZÁVĚR

Úkolem bylo vytvořit snadno použitelné, stabilní a zároveň velice obsáhlé grafické uživatelské rozhraní pro práci s datovými archivy. To je umožněno díky tomu, že funkčnost uživatelského rozhraní není pevně dána a je možné ji podle potřeb upravovat či rozšiřovat pomocí sad algoritmů. Zároveň není uživatelské rozhraní omezeno na konkrétní typy dat v archivu, jelikož je schopné samo detekovat a zobrazit čistý text, formátovaný text, multimediální soubory a jiné. Pro projekt JMZW má grafické uživatelské rozhraní přínos v tom, že je možné provádět dohled nad sběrem a čištěním dat. Také je velmi usnadněno vytváření nových algoritmů či interaktivní sestavování jazykového modelu.

**Poděkování:** Příspěvek byl podpořen grantovým projektem Západočeské univerzity v Plzni, projekt č. SGS-2010-054, a za podpory Ministerstva průmyslu a obchodu, projekt č. MPO FR-TI1/486.

## REFERENCE

Jan Švec, Jan Hoidekr, Daniel Soutner, Jan Vavruška, 2011. *Web Text Data Mining for Building Large Scale Language Modelling Corpus*.

Qt Online Reference Documentation <http://www.python.org/doc/>.

Python Documentation <http://doc.qt.nokia.com/>.