

НАУЧЕН ОТЧЕТ ЗА ВТОРИЯ ЕТАП

на изпълнението на договор ДТК-02-54/2009 „Информационни технологии за представяне с музика, ноти и текст на български народни песни в дигитална библиотека”

от колектива на

Факултет по математика и информатика – Софийски университет „Св. Климент Охридски”

1. Изпълнение на работната програма на колектива от ФМИ – СУ за втори етап

РП4. НАУЧНИ ИЗСЛЕДВАНИЯ В ИТ И ДБ

4.2. Семантично ориентирано търсене в ДБ

Развита е разработената на първия етап от изпълнението на проекта методика за реализация на семантично търсене в дигитална библиотека, което съответства на изискванията и очакванията на потенциалните потребители на библиотеката – експерти в съответната предметна област. Въведено е понятието “*образец на понятие*“ (*concept search pattern*) – образец на характерно стилистично средство, който може да се съпостави с относително голяма част от текста на песен и има определено значение като израз например на любов, омраза, ревност, изневяра и т.н. Определено е множеството от специални символи, които могат да се използват в образците, заедно с правилата за тяхното успешно съпоставяне със съответни части от текстовете.

Предложен е подход за анализ на настроението в текстовете (лириката) на песните – на отделна песен или определена група песни, например множество песни, резултат от изпълнението на дадена заявка за търсене. Този подход може да бъде характеризиран като обобщение на популярния *lexicon-based sentiment analysis*. За целта с някои от класовете от предметната онтология са асоциирани подходящи положителни или отрицателни числа, които играят ролята на оценки на настроението, носено от съответните понятия. Настроението на дадена песен се определя в съответствие със сумата от оценките на настроенията на отделните думи в текста на песента. По-специфичните понятия от онтологията и всички техни форми и синоними наследяват оценките на настроението на съответните по-обща понятия.

4.4. Инструменти за семантично търсене

Създадена е разширена версия на машината за търсене в дигитална библиотека с български фолклорни песни, чието хранилище включва: текстове (лирика) на песните; нотни текстове на песните; музикални (MP3) файлове с автентични

изпълнения на песните; музикални (MIDI) файлове, генерирани с помощта на LilyPond от нотните текстове на песните (*Приложение ФМПИ*). В новата версия на машината за търсене са обогатени средствата за семантично търсене, като за целта е разширена предметната онтология, описани са множество образци на понятия и са реализирани възможности за търсене и съпоставяне на части от текстовете на песните с образци на понятия.

Създадена е и експериментална версия на софтуерен инструмент за анализ на настроението в текстовете (лириката) на песните – на отделна песен или определена група песни, например множество песни, резултат от изпълнението на дадена заявка за търсене.

Новата версия на машината за търсене е реализирана с използване на технологията ASP.NET 3.5, език за програмиране C# и IDE Visual Studio 2008. Предметната онтология е създадена с помощта на популярния редактор на онтологии Protégé-OWL 4.1.

РП8. РАЗПРОСТРАНЕНИЕ НА РЕЗУЛТАТИТЕ ОТ ПРОЕКТА

8.2. Разпространение в научни институции

Получените резултати от работата по проекта от страна на колектива на ФМИ – СУ са представени на :

- Петнадесетата международна конференция по електронно публикуване ELPUB 2011 “Digital Publishing and Mobile Technologies” (Иstanbul, Турция, 22-24 юни 2011 г.);
- Шестнадесетата международна конференция по електронно публикуване ELPUB 2012 “ Social Shaping of Digital Publishing: Exploring the Interplay between Culture and Technology” (Гимараеш, Португалия, 14-15 юни 2012 г.);
- Шестата международна конференция по информационни системи и грид технологии ISGT 2012 (София, 1-3 юни 2012 г.);
- Седмата международна конференция по информационни системи и грид технологии ISGT 2013 (София, 31 май-1 юни 2013 г.).

8.5. Публикации

Получените резултати от страна на колектива на ФМИ – СУ са отразени в четири публикации, посочени в т. 2 на настоящия отчет.

2. Списък на публикациите

Резултатите от работата по проекта на членове на колектива от ФМИ – СУ са отразени в следните публикации:

- Nisheva-Pavlova, M., P. Pavlov. Semantic Search in a Digital Library with Bulgarian Folk Songs. In: Y. Tonta et al. (Eds.), “Digital Publishing and Mobile Technologies. 15th International Conference on Electronic Publishing, June 22-24, 2011, Istanbul, Turkey”, Hacettepe University, Ankara, 2011, ISBN 978-975-491-320-0, pp. 103-109 – Приложение **ФМИ2**¹;
- Armyanov, P., R. Goranova. GUI Tools for Complex Queries in RDBMS. Proceedings of the Fifth International Conference on Information Systems and Grid Technologies (Sofia, May 27-28, 2011), ISSN 1314-4855, Sofia, St. Kliment Ohridski University Press, 2011, pp. 75-81 – Приложение **ФМИ3**;
- Nisheva-Pavlova, M., P. Pavlov. Ontology-Based Search and Document Retrieval in a Digital Library with Folk Songs. Information Services and Use, ISSN 1875-8789, Vol. 31, Numbers 3-4 (2011), pp. 157-166, DOI 10.3233/ISU-2012-0645 – Приложение **ФМИ4**;
- Nisheva-Pavlova, M., P. Pavlov, D. Shukerov. Semantic Digital Library with Bulgarian Folk Songs. Proceedings of the Seventh International Conference on Information Systems and Grid Technologies (Sofia, May 31-June 1, 2013), ISSN 1314-4855, Sofia, St. Kliment Ohridski University Press, 2013 (to appear) – Приложение **ФМИ5**.

¹ Тази публикация е включена като приета за печат и в научния отчет за първия етап на изпълнението на договора.