

ОБОБЩЕН НАУЧЕН ОТЧЕТ

за 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”**

Основни научни резултати

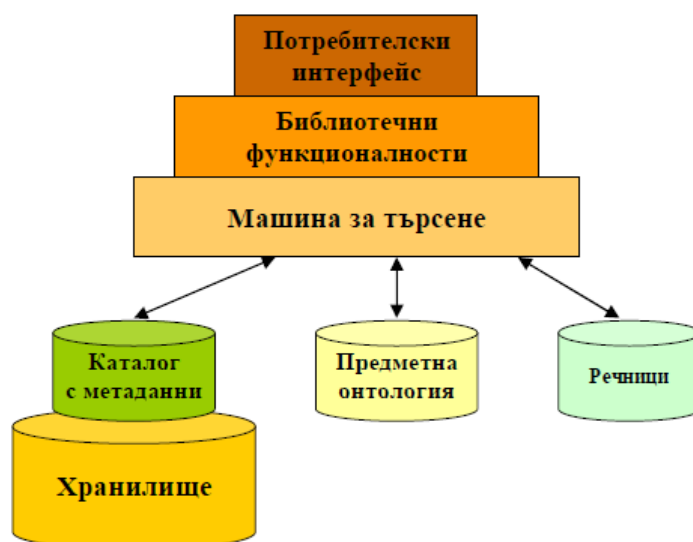
Основна цел на проекта е създаването на методи и софтуерни средства, които да отговорят на информационно-технологичните потребности на изследователите на музикалния фолклор в областта на етнологията, етномузикологията и фолклористиката. В изпълнение на тази цел от колектива от ФМИ – СУ са извършени изследвания, насочени към:

- анализ на особеностите на областта и създаване на функционален модел на дигитална библиотека с българска фолклорна музика;
- развитие и прилагане на информационни технологии за създаване на дигитални библиотеки, методи и инструменти за търсене (в частност, семантично търсене) в хетерогенна дигитална библиотека с материали на различна тематика.

Резултатите от изследванията са приложени при създаването на прототип на дигитална библиотека (наречена DjDL) с български фолклорни песни, представени с техните нотни записи, текст и музикални изпълнения, която предоставя разнообразни инструменти за достъп до съхраняваното хетерогенно дигитално съдържание.

Понастоящем в DjDL се съхранява колекция от съставни (многокомпонентни) дигитални обекти, които представят част от непубликувания архив с ръкописни теренни записи на проф. Тодор Джиджев, дигитализирани от колективите от ИМИ, ИЕФЕМ и ИИИ – БАН.

Функционалната структура на DjDL е показана на Фигура 1. Тя следва принципите на референтния модел за дигиталните библиотеки (Digital Library Reference Model, DLRM), разработен от няколко европейски изследователски групи в рамките на DELOS Network of Excellence on Digital Libraries.



Фигура 1. Функционална структура на DjDL

Каталогът с метаданни съдържа кратки описания в XML формат на песните, включени в хранилището на библиотеката. Типични примери за метаданни на фолклорните песни са: име на песента, жанрова класификация според различни класификационни признаци, изпълнител(и), записвач(и), дешифратор(и), място и дата на записа и др.

Хранилището на DjDL включва съставни дигитални обекти, които представят български фолклорни песни. Отделните компоненти на тези обекти съдържат съответно:

- текстовете (лириката) на песните;
- нотните текстове на песните в текстовия формат, поддържан от системата LilyPond;
- музикални (MP3) записи с автентични изпълнения на песните;
- записи (в MIDI формат) на мелодиите на песните, генерирани с помощта на LilyPond от нотните текстове.

Предметната онтология представлява съвкупност от няколко свързани (под)онтологии, които играят ключова роля за реализацията на проектираните функционални възможности на машината за търсене. Те описват множество понятия от различни тематични области (бит и семейство, обществени отношения, исторически събития, природни състояния и явления, астрономически явления и мн. др.), свързани със съдържанието на песните, както и свойства на понятията и релации между тях.

Текущата версия на машината за търсене на DjDL предоставя възможности за търсене само в явно формулираните метаданни и текстовете (лириката) на песните. Тя поддържа два основни типа търсене: търсене по ключови думи/фрази и семантично търсене. При семантичното търсене освен предметната онтология се използват съществено и експериментални версии на два малки по обем речника в

дигитален (XML) формат, създадени за целите на тестването на машината за търсене – речник на остарели и диалектни думи и синонимен речник.

Множеството от метаданни за фолклорните песни е определено, като са взети предвид международно утвърдени стандарти (Text Encoding Initiative (TEI) и Encoded Archival Description (EAD)) и класификационни схеми, както и особеностите на националната традиция. В съответствие със стандарта EAD всяко каталожно описание съдържа метаданните за определена фолклорна песен и текста (лириката) на тази песен. По такъв начин се постига по-висока скорост на работата на машината за търсене.

Фигура 2 съдържа списък на елементите (предвидените типове метаданни и текстът на песента, записан като стойност на елемента <folk_song_text>), от които са съставени каталожните описания на песните от хранилището на DjDL, и кратко описание на значението на всеки от тях.

**Структура на XML документ за описание на народна песен,
включващо и текста на песента**

<folk_song_desc>	главен елемент, на най-високо ниво
<folk_song_title>	име на народната песен
<classification>	класификация, може да са няколко такива елемента
<folk_song_text>	текст на народната песен, форматиран като стихотворение
<singers>	певци, може да е един певец с името му; може да са няколко певци с имената им поотделно; може да са няколко певци, но в група отделен певец с данни за него или група певци с известните данни за тях
<singer>	
<sing>	име на певица или информацията за група певци с известните данни за тях
<gender>	пол на певица (мъж/жена)
<bornYear>	година на раждане на певица
<bornPlace>	име на населеното място - месторождение на певица
<bornArea>	име на окръг/околия, в който/която се намира месторождението на певица
<immigration>	данни за преселничество на родителите на певица
<marriageYear>	година на омъжване/женитба
<marriagePlace>	име на населено място, в което е омъжена певицата
<marriageArea>	име на окръг/околия
<livePlace>	име на населено място, в което живее певецът
<liveArea>	име на окръг/околия
<writers>	записвач(и) (изследовател(и), записал(и) текста на песента)
<writer>	пълно име на записвач, както е известно
<decipherers>	дешифратор(и)
<decipherer>	пълно име на дешифратор, както е известно
<record>	дата, на която е направен записът (може да е само годината)
<day>	ден (част от датата на записа)
<month>	месец (част от датата на записа)
<year>	година (част от датата на записа)
<filetext>	име на файла (.txt) с текста на песента
<filenotes>	име на файла (.ly) с нотите на песента
<filemusic>	име на файла (.mp3) със запис на изпълнението на песента
<filemidi>	име на файла (.mid) с генерирания от нотния текст звук на песента
<note>	бележка/текст на записвача за поясняване на обстоятелства
<dict>	речник на непознати за записвача думи
<term>	непознат термин
<def>	значение, обяснение

Фигура 2. Структура на каталожно описание на фолклорна песен в DjDL

Предметната онтология описва определено количество знания от няколко области, с които е свързана лириката на българските народни песни (с дефиниции на основни понятия от тези области, техни свойства/релации и подходящи представители/екземпляри на някои понятия). Тя се състои от множество свързани с общи свойства (под)онтологии, разработени специално за случая, които са

необходими за работата на машината за търсене. Списъкът на по-съществените елементи на това множество включва:

- онтология на фолклорните песни – представя различни жанрови класификации на фолклорните песни (например по контекст на изпълнение – трудови, празнични и др.; по културни функции – балади, благословии, клетви, инструменти на флирта и т.н.);
- онтология на тема “бит и семейство” (облекло, занятия, пособия, типични места, роднински отношения, семейни събития, празници, традиции и ритуали и др.);
- онтология на тема “природни състояния и явления”;
- онтология на тема “астрономически явления”;
- онтология на тема “обществени отношения и събития в обществения живот” (размени/сделки, избори, размирици и др.);
- онтология на тема “исторически събития”;
- онтология на митични същества и демони;
- “географска” онтология (общини/околии и населени места в тях с географски координати – според административното деление на страната в началото на втората половина на 20-ти век и понастоящем).

Предметната онтология е разработена с помощта на Protégé-OWL. Повечето класове от онтологията са конструирани като дефинирани OWL класове чрез необходими и достатъчни условия, формулирани в термините на подходящи ограничения върху определени свойства.

Свойствата “form” (“форма”) и “synonym” (“синоним”) предоставят на машината за търсене различни граматични форми и синоними на термините, използвани като имена на класове от предметната онтология.

В лириката на фолклорните песни се срещат голям брой сравнения, метафори, олицетворения, алегии и други стилистични фигури. Затова е целесъобразно машината за търсене на DjDL да интегрира в процеса на семантично търсене подходяща методика за използване на наличната предметна онтология с други техники на Изкуствения интелект.

В този смисъл са дефинирани множество образци на характерни стилистични средства, които могат да се съпоставят с относително големи части от текстовете на песните и имат определено значение като израз например на любов, омраза, ревност, изневяра и др. Наричаме ги *образци на понятия (concept search patterns)*. 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.

Машината за търсене на DjDL поддържа два основни типа търсене: търсене по ключови думи/фрази и семантично търсене. Нейната текуща версия предоставя възможности за търсене само в каталожните метаданни и текстовете (лириката) на песните.

Потребителската заявка за търсене по ключови думи съдържа произволен брой (една или повече) думи/фрази в текста или метаданните – потребителят определя връзката (конюнкция или дизюнкция) между тях. Потребителят посочва и в кои източници (в текстовете на песните, в определени метаданни от каталожните описания или комбинирано в текстовете на песните и метаданните) да се

осъществи търсенето по отношение на всеки елемент на заявката. Извежда се списък от хипервръзки към намерените файлове с каталожни описания и текстове на песни, който може да бъде сортиран по различни критерии според решението на съответния потребител (по имената на песните, по имената на изпълнителите и др.). След щракване с мишката върху името на избран файл се визуализира текстът на съответната песен. Потребителят има възможност да получи и достъп до съхраняваните музикални (MP3 и MIDI) записи на песните.

Модулът, който осъществява семантичното търсене, понастоящем изпълнява само “атомарни” заявки, всяка от които е съставена от една дума или фраза, зададена от съответния потребител. Потребителят формулира заявката и посочва източниците (текстовете на песните или конкретни метаданни), в които ще се извърши търсенето. В процеса на семантично търсене се извършва подходящо разширяване на потребителската заявка, като за целта се използват наличните явно формулирани знания за предметната област (предметната онтология и речниците). На потребителя се предоставя възможност да прецизира получената разширена заявка. При възможност се използват и образци на понятия, приложими в конкретния случай.

Най-съществен източник на знания за разширяване на потребителската заявка е таксономията (“is-a” йерархията) на понятията/класовете, включена в предметната онтология. Най-напред в процеса на разширяване на заявката се извършва обхождане в широчина на графа, който представя “is-a” йерархията на класовете, стартирайки от възела, съответен на класа/понятието от оригинално формулираната заявка. Имената на посетените възли, които фактически са по-специфичните понятия от това, интересуващо потребителя (частните случаи на това понятие), се добавят към разширената заявка, като се избягват евентуални дублирания. Полученият списък от понятия се визуализира по подходящ начин и се предоставя на потребителя за евентуално “ръчно” прецизиране.

По време на следващата стъпка от процеса на разширяване към потребителската заявка се добавят и някои форми и синоними на натрупаните до момента термини, които са стойности на свойствата “форма” и “синоним” на съответните класове или се съдържат в наличните речници. После се добавят и имената на всички съществуващи екземпляри на класовете от разширената заявка. Накрая се добавят и стойностите на такива свойства на екземплярите, включени в заявката на предишната стъпка, които са специфицирани явно като съществени за съответните класове по отношение на семантичното търсене (такова е например свойството “участници” за екземплярите на класа “историческо събитие”).

Така оригиналната заявка за търсене, формулирана от съответния потребител, е разширена в максималната възможна степен и фактически представлява дизюнкция на всички включени термини и имена на техни екземпляри (представители). Разширената заявка се предоставя на потребителя за евентуално повторно прецизиране, след което се преминава към изпълнението ѝ. В случаите, когато търсенето се осъществява в текстовете на песните и сред понятията от разширената заявка има такива, за които са дефинирани съответни образци, модулът за съпоставяне по образец претърсва допълнително текстовете с цел проверка за съпоставимост с всеки от тези образци.

Машината за търсене на DjDL поддържа също възможности за изпълнение на заявки, които предполагат проверка за равенство или определен тип неравенство.

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. 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 настроеността 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.

Като най-значим резултат от изследванията, осъществени от научния колектив от ФМИ – СУ, може да бъде посочен предложеният подход за реализация на семантично търсене в дигитална библиотека, което съответства на изискванията и очакванията на потенциалните потребители на библиотеката – експерти в съответната предметна област.

Понастоящем се провеждат експерименти с цел получаване на аргументирана оценка на функционирането на машината за търсене на DjDL. В частност, работи се по установяване на средната точност на получаваните резултати при семантично търсене.

Следваща стъпка ще бъде обогатяването на машината за търсене със средства за семантично търсене и откриване на закономерности в нотните текстове на песните. Основна цел в това отношение е създаването на инструмент за автоматизиране на изследванията на някои музикални характеристики на българските фолклорни песни (например техните мелодии и ритми) и откриване на сходства между тях по отношение на различни критерии.

По такъв начин DjDL ще разполага с множество от разнообразни софтуерни инструменти, подпомагащи изследванията в области като фолклористика, филология и музикология.

Резултати от работата по проекта на членове на колектива от ФМИ – СУ са представени на шест международни научни конференции. Общо пет статии са отпечатани или приети за печат в международни списания или сборници на международни конференции. Работата по проекта допринесе за успешното участие на доц. д-р Мария Нишева-Павлова в конкурс за заемане на академичната длъжност “професор” във ФМИ – СУ.

Списък на публикациите

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

1. Peycheva, L., N. Kirov, M. Nisheva-Pavlova. Information Technologies for Presentation of Bulgarian Folk Songs with Music, Notes and Text in a Digital Library. Proceedings of the Fourth International Conference on Information

- Systems and Grid Technologies (Sofia, May 28-29, 2010), St. Kliment Ohridski University Press, 2010, ISBN 978-954-07-3168-1, pp. 218-224;
2. 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;
 3. 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;
 4. 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;
 5. 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).

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

- Четвъртата международна конференция по информационни системи и гريد технологии ISGT 2010 (София, 28-29 май 2010 г.);
- Пролетната научна сесия на ФМИ (София, 26 март 2011 г.);
- Петата международна конференция по информационни системи и гريد технологии ISGT 2011 (София, 27-28 май 2011 г.);
- Петнадесетата международна конференция по електронно публикуване ELPUB 2011 "Digital Publishing and Mobile Technologies" (Истанбул, 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 гريد технологии ISGT 2012 (София, 1-3 юни 2012 г.);
- Седмата международна конференция по информационни системи и гريد технологии ISGT 2013 (София, 31 май-1 юни 2013 г.).