

МАТЕМАТИКА И МАТЕМАТИЧЕСКО ОБРАЗОВАНИЕ, 2000
MATHEMATICS AND EDUCATION IN MATHEMATICS, 2000
*Proceedings of Twenty Ninth Spring Conference of
the Union of Bulgarian Mathematicians
Lovetch, April 3–6, 2000*

**ЗНАЕ ЛИ СВЕТАТ, ЧЕ И НИЕ НЕЩО СМЕ МУ ДАЛИ?
СТАРОВЪЛГАРСКИТЕ РЪКОПИСИ ПРЕЗ ПРИЗМАТА НА
ИНФОРМАТИКАТА**

Милена Добрева

В доклада е направен обзор на приложенията на информационните технологии в областта на представянето и изучаването на старобългарски ръкописи. В него са разгледани накратко досегашните постижения на базите от данни, представянето и обработката на текстове и изображения. Направен е изводът, че е назряла нуждата от разработване на материали за дистанционно обучение в областта на старобългаристиката.

Как се е променял старобългарският език, за да стигне до езика, на който говорим днес? А какво е станало с него в руските или сръбските земи, за да зазвучи там реч, която толкова се различава от нашата? Защо на различни места езиковите изменения водят до толкова различни резултати? Колко време е нужно, за да се промени нещо в езика? Какво се променя повече – граматичните конструкции или думите, които се употребяват? Тези въпроси нямат лесен и бърз отговор. Десетилетия наред езиковедите изучават средновековните славянски ръкописи, като основният им „инструмент“ е сравняването на различни писмени паметници.

Подобни сравнения са много трудоемки. Представете си, че трябва да препишете един текст, като всяка отделна дума заедно с изречението, в което се появява, се пише на отделен картотечен фиш. . . Добавете към това различна помощна информация – кой е текстът, на коя позиция в него (лист и ред) се появява съответната дума: Подредете след това всички подобни фишове по азбучен ред. . . Направете това за друг текст и се опитайте да сравните например какви думи се появяват и в двата текста, и колко пъти става това. . . Ясно е, че един текст на съвременен вестник би се преписвал по този начин повече от седмица. А ако това трябва да се направи за ръкопис от 200 листа? След като разполагаме със съвременните информационни технологии, трябва ли да работим по познатите начини?

В този доклад ще представя накратко досегашния опит за приложения на информационните технологии при представянето и изучаването на старобългарски ръкописи. Основните области на информатиката, които намират приложение в подобни начинания, са *базите от данни* и *представянето и обработката на текстове и изображения*. Ще се спра на някои възможни приложения от областта на дистанционното обучение.

В началото бе...

Създаване на електронни каталози на ръкописи. Преди около петнадесет години някои специалисти от областта на старобългаристиката са започнали да търсят начини да облекчат своята работа с помощта на информатиката. Първият подобен опит е направен в Холандия, където група от специалисти по средновековни текстове с помощта на информатици са разработили система за описание на средновековни славянски ръкописи [3].

Ориентирането към подобна област на приложение на информатиката съвсем не е случайно – славянското средновековно ръкописно наследство наброява стотици хиляди ръкописи и ръкописни фрагменти, пръснати в хранилища в различни страни. Точният брой на средновековните славянски паметници все още не е известен. Само в български хранилища се пазят над 8500 средновековни ръкописа, голяма част от които са на старобългарски език. Тъй като славянски ръкописи има в множествобирки из цяла Европа и бившия Съветски съюз и липсват дори печатни каталози за част от тях, все още няма точна картина на средновековното ни ръкописно наследство.

Опитите за създаване на компютърни каталози на сбирките целят да се подобри достъпът до данните за налични ръкописи – както в рамките на една ръкописна сбирка, така и в рамките на множество от библиотеки. Но и до днес, след около петнадесет години практическа работа в това направление, нерешените въпроси са повече от решените.

Основният проблем в тази област се дължи на липсата на общоприет модел какви данни за ръкописите трябва да се представят в един електронен каталог – броят на включените в описанията данни може да варира силно според възгледите на специалистите старобългаристи, които участват в разработването на проекта. Обикновено подобни описания са много сложни и могат да съдържат около 120 различни елемента, които освен библиографското описание на паметника включват и представяне на различни негови особености – например какво е съдържанието му (в каква последователност и какви текстове включва конкретният ръкопис), кратка характеристика на писмото, езиковите особености, украсата и др.

Напоследък се проявява още едно затруднение при създаването на електронни описания на ръкописи. Традиционно се е предполагало старобългаристите да познават достатъчно добре старобългарски и съвременен български език. При съвременните тенденции за постепенно натрупване на електронни ресурси, пръснати в различни страни и възможността за публикуване на подобни данни в Интернет, се разширява кръгът на хората, които потенциално биха получили достъп до електронни издания, свързани с нашето културно наследство. По тази причина представянето на данни за славянските ръкописи започва да се ориентира към използването на английски език – тенденция, която допълнително затруднява процеса на въвеждане на данни за ръкописи, достатъчно сложен и трудоемък сам по себе си [2]. Стига се до положението основните изследователски данни за наши ръкописи да се въвеждат изключително на английски език с изключение на текстовите фрагменти, включени

в описанието (виж например проекта, представен в [6]).

Един интересен опит в това направление на работа е включването на *Ръкописния отдел на Националната библиотека „Св. Кирил и Св. Методий“* като асоцииран член на проекта MASTER (Manuscript Access through Standards for Electronic Records – Достъп до ръкописи чрез стандарти за електронен запис), който е финансиран от Европейската общност и включва организации от Великобритания, Дания, Италия, Франция, Холандия и Чехия [5]. Въвеждането на данни става с помощта на специално разработена дефиниция за тип документ (DTD) като специализирана надстройка на стандарта на Международната организация за стандартизация за маркиране на текстове [4]. Участниците в това начинание работят с амбицията да създадат световен стандарт в областта на електронното описание на ръкописи.

В рамките на работата по този проект у нас се експериментира в две направления: от една страна, описанията се правят успоредно на два езика – български и английски. От друга страна, се въвеждат данни за ръкописи, които не са написани на латиница и принадлежат към славянската писмена традиция, която съществено се различава от западната.

Представяне и анализ на текстове. Следващата област от информационните технологии, която съвсем естествено е намерила място в работата на старобългаристите и другите специалисти по средновековна славистика, е текстообработката в най-широк смисъл. Навлизането на компютърните текстообработващи системи е било съпроводено от разработването на специализирани шрифтове за писане на глаголица и средновековна кирилица. Но и досега в тази област няма общоприет стандарт за представяне на тези азбуки, което затруднява електронната размяна на текстове между специалисти от различни институции.

Проблемът е в това, че за разлика от повечето съвременни езици, където азбучният състав и изписванията на буквите са стандартизирани, средновековните славянски текстове могат сериозно да се различават по буквен състав (в някои паметници се употребяват около 30 букви, докато в други – над 40). В множество случаи е трудно да се определи дали разликата в изписването на определени букви е резултат от невъзможността на ръка да се изпишат абсолютно еднакви знакове, или има специално значение за разчитането на текста.

Така разяснението на Черноризец Храбър защо се е наложило да се създава писмо специално за славяните, чиито думи не могат да бъдат винаги добре представени чрез гръцката азбука, в компютърната ера се измества така: дава ли един конкретен начин за кодиране на старите ни текстове възможност да се представят старобългарските паметници в цялото им разнообразие? Най-изчерпателното до момента предложение за компютърно кодиране на старата кирилица е направено в [1], където са включени 205 знака.

Тъй като задачите, които старобългаристите решават, са изключително разнообразни, не съвпадат и подходите на специалисти от различни школи към представянето на средновековните ни текстове в компютърен вид. В някои случаи интерес



Фиг. 1. Фрагмент от факсимиле на ръкопис, XVI в.

представяват всички подробности от разположението на текста – кои букви са изписани над основния или под основния писмен ред, ако има няколко начертания на една и съща буква, къде се появяват те, как точно са съкращавани думите в текста (освен буквения състав, и съкращенията в старобългарските текстове не са нормирани). На Фиг. 1 е представен един сравнително късен наш ръкопис, в който могат да се видят изброените току-що особености.

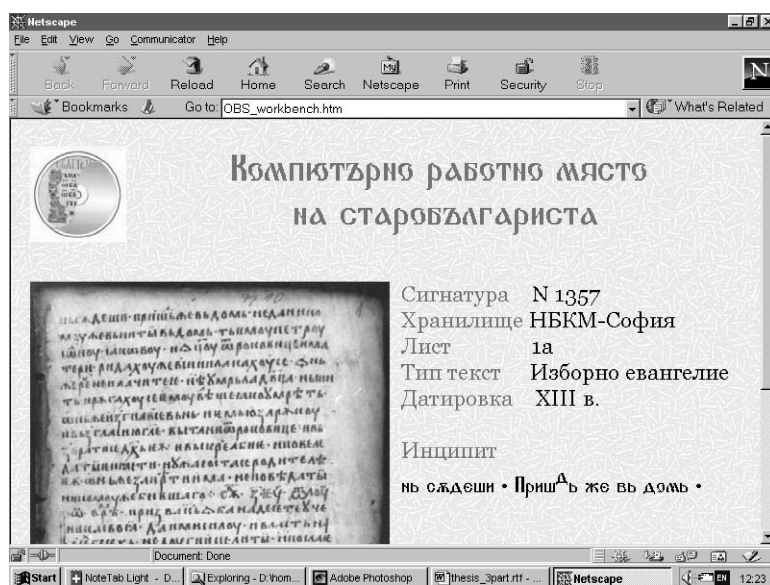
Текстовите редактори не позволяват въвеждането на подобни нетипични за съвременните текстове елементи, нито съществуват средства за търсене в текстове, в които всяка дума може да се появи съкратена и разположена по множество различни начини. В някои случаи обаче точното разположение на всички текстови елементи не се смята за толкова съществено, и тогава текстовете се въвеждат в т.нар. *нормализиран вид* – правописът е уеднаквен, а съкратените думи се изписват изцяло. Този подход несъмнено облекчава въвеждането на текстове и автоматизираното търсене в тях, но при него се губят важни данни, които са показателни за развитието на

писмената култура в Средновековието.

Поради тези затруднения с представянето на старобългарски текстове досега почти няма примери на използване на компютрите за нещо повече от рутинна текстообработка и предпечатна подготовка. Така съпоставителните изследвания на различни текстове се правят все чрез „ръчно“ фиширане и сортиране.

Представяне и обработка на изображения. Особеностите на средновековните ръкописни текстове, представени по-горе, показват защо представянето само на текста на даден паметник не може да даде достатъчно пълна представа за него. Представянето на изображенията може да послужи и за подобряване на четливостта на текста в ръкописи, които са пострадали от влага или при пожар.

Възможността да се изучава текстът на даден ръкопис в съчетание с точни изображения на паметника е от изключителна важност за всички, които се интересуват от средновековните ни текстове. Така всъщност стигаме до един от закономерните парадокси на приложенията на информационните технологии в областта на средновековното ръкописно наследство: средновековният ръкопис представлява един цялостен обект от гледна точка на изследователя. Компютърното представяне на този обект обаче го „разлага“ на изображения, текст и структурирани данни за него, като всяко различно представяне се използва за специфични цели.



Фиг. 2. Част от описание на ръкопис, подготвено за публикуване в Интернет

Съвременните мултимедийни технологии помагат да се обединят, поне на механично ниво, подобни различни представяния. На Фиг. 2 е представена една идея как може да се съчетае изображението на един лист от ръкописен фрагмент с кратки

данни за паметника, като са използвани средствата на съвременните технологии за публикуване в Интернет.

Вместо заключение: някои приложения в обучението. Средновековните ни ръкописни текстове имат отношение към широк кръг от дисциплини, изучавани в българското училище. Работата с текстове, написани на старобългарски език в оригинал обаче е рядкост. Учениците на практика не се запознават с тази част от нашето културно наследство освен чрез някои илюстрации и ползване на старобългарските текстове в превод на новобългарски език.

При това положение по-подробното запознаване със старобългарските паметници остава за тези ученици, които имат специални интереси в областта. Но дори и те не могат да разчитат на директен достъп до самите паметници в хранилищата.

В такива случаи е разумно да се използват възможностите на дистанционно обучение. За жалост в област, която е неразривна част от нашето историческо наследство, все още няма никакви достъпни ресурси. Малкото, което може да се намери в Интернет за нашата култура, е написано предимно на английски език. Източниците, които са ползвани, не винаги отразяват достатъчно пълно разглежданите проблеми. Преобладаващият брой интернетски „страници“, представящи аспекти на нашата средновековна култура, са изготвени в чужбина. При това положение не е чудно човек да се сблъска с бисери от типа на това, че „България е известна по света със своята шопска салата“. Как да не се запитаме в подобни случаи, ако и ние сме дали нещо на света, той знае ли за това – поне според отражението на нашето културно наследство в световната информационна мрежа. . .

Разработването на електронни публикации за старобългарската писмена култура не е лека задача. Предаването на огромен по обем материал е свързано не само със значителна по обем техническа работа, но и със създаването на подходящи модели за представяне на средновековната ни култура. Не ми е известно у нас да са правени опити за публикуване на материали за обучение на ученици и студенти, освен в рамките на проекта „Делото на Кирил и Методий: славяните и Византия през IX в.“, който е в процес на разработване в сътрудничество на ИМИ с Американския Университет в България. Целта на този проект е да се подготвят учебни материали, които да дават начална представа на чуждестранни студенти в областта на славистиката и българистиката за възникването и разпространението на славянската писменост.

По-горе очертах накратко областите, в които в последните петнадесет години у нас се е натрупал известен опит. Време е този опит да се обедини и да послужи за едно достойно представяне на културното ни наследство, включително и за средновековните ръкописни паметници, в Интернет. В случая не е важно само да представим на света своето лице, но и да проявим грижа за това какво научават нашите деца за своята страна.

ЛИТЕРАТУРА

- [1] D. BIRNBAUM. Standardizing Characters, Glyphs and SGML Entities for Encoding Early Cyrillic Writing. *Computer Standards and Interfaces*, **18** (1996), 201-252.
- [2] M. DOBREVA. Use of SGML by Philologists. In: Proceedings of SGML-Belux conference, Brussels, October 30-31, 39-53.
- [3] A. J. GEURTS, A. GRUIJS, J. VAN KRIEKEN, W. R. VEDER. Codicography and Computer, In: *Палата книгописная*. **17-18** (1987), 4-29.
- [4] International Organization for Standardization, ISO 8879: Information processing-Text and office systems-Standard Generalized Markup Language (SGML), Geneva, ISO, 1986.
- [5] <http://www.cta.dmu.ac.uk/projects/master>
- [6] A. MILTENOVA. Computer Repertory of Medieval Literature and Letters. In: M. Dobрева (ed.) *Text Variety in the Witnesses of Medieval Texts*, Proceedings of the Int. Workshop, Sofia, 21-23 September 1997, S., 1998, 138-149.

Милена Добрева
Институт по математика и информатика
Българска академия на науките
ул. „Акад. Г. Бончев“ бл. 8
1113 София
e-mail: dobрева@math.bas.bg

IT APPLICATIONS TO MEDIAEVAL BULGARIAN MANUSCRIPTS' STUDIES

Milena Dobрева

The paper presents a survey of the applications of information technology to the field of presentation and study of mediaeval Bulgarian manuscripts. It gives a brief summary of the development of different applications – DBMS, text and image presentation and processing. The necessity of developing distance learning materials in the field of Mediaeval Bulgarian studies is discussed. Creating formal models, developing proper encoding and presenting material in different languages are amongst the key problems in this field.