

французских текстов. Одной из целей консорциума является создание единого портала, позволяющего осуществлять поиск во всех базах данных его участников. Совместными усилиями решаются проблемы морфологической разметки и лемматизации текстов, ведется исследование типологии текстов, призванное обеспечить более адекватную оценку репрезентативности существующих корпусов и наметить приоритетные направления их расширения.

Список литературы

- Heiden S., Guillot C. Capitalisation des savoirs par le web : une application de la TEI pour l'encodage et l'exploitation des textes de la Base de Français Médiéval // Ancien et moyen français sur le Web, enjeux méthodologiques et analyse du discours / Kunstmann, P. (ред.). — Ottawa: Les éditions David, 2002. — С. 77–92.
- Heiden, S., Lavrentiev, A. Ressources électroniques pour l'étude des textes médiévaux : approches et outils // Revue française de linguistique appliquée. — 2004. — IX(1). — С. 99–118.
- Kunstmann, P. Ancien et moyen français sur le Web: textes et bases de données // Revue de linguistique romane. — 2000. — 253–254. — С. 17–42.

Проект INCA: цели, проблемы, тенденции и ИТ-решения по капитализации знаний

С. Г. Маслов

Ижевский государственный технический университет, Россия

The project on the knowledge capitalization system in the form of constructive electronic publication fund for technical as well as natural-scientific educational and research activities is considered. The project objectives and tasks are denoted. A realization approach based on the original component basis and synthesis of the system, cybernetic and synergetic approaches is suggested

Бурное развитие человечества происходило в периоды появления новых форм представления и методов переработки информации и знаний, а также синтеза существующих технологий. На современном этапе такой формой становятся электронные издания, накапливаемые в электронных библиотеках, на сайтах, на порталах и в электронных каталогах. Однако сложившегося разнообразия явно недостаточно, потому что оно концентрирует исследователя на уже известных ему формах, но существующих на других носителях. Это чаще всего приводит к избытку некачественной информации или дезинформации, к ситуации, когда излишек информации вреден точно так же, как и ее недостаток.

В работе рассматриваются современные проблемы, тенденции и ИТ-решения, которые инициировали *проект INCA* (фонд конструктивных электронных изданий в области инженерно-технического образования и научно-исследовательской деятельности) и составили основу его концептуальных архитектурных решений.

Фонд конструктивных электронных изданий — это стратегическое направление поддержки общества, основанного на знаниях, на базе непрерывности потока информации и знаний в цепи: *исследование и творчество* (анализ, генерация и синтез знаний); *обращение* (распространение информации и знаний, формирование

и саморазвитие интеллектуальных ресурсов); *инновации* (производство и потребление на основе знаний).

Цель проекта — создание инвариантных и масштабируемых решений по построению конструктивных электронных изданий и фонда электронных изданий, отражающих систему знаний о предметной области, об инструментах, о субъектах, ресурсах и процессах через различные аспекты содержания, логической структуры, формы и действий на основе конструктивных, мультимедийных и полисенсорных представлений.

Некоторые количественные характеристики научных и технических целей: семантическое сжатие информации до 3 раз за счет формализованного описания действительности согласованными полисенсорными статическими и динамическими представлениями; повышение качества, эффективности и скорости восприятия информации от 2 до 5 раз за счет адаптации к читателю, сокращения объема воспринимаемой информации и переключения на процессы построения информации. Эффективность фонда планируется для субъектов с численностью от 1 до 1000 человек.

Конечный результат проекта — разработка фонда электронных изданий с первичным наполнением для инженерно-технического и естественнонаучного образования и исследования, построенного на конструктивной компонентной базе, создаваемой на основе междисциплинарного сотрудничества. Начальная версия фонда должна показать функциональные возможности и преимущества принципиально нового класса электронных изданий и методов взаимодействия с ним.

Фонд предназначен для концентрации научных знаний, которые позволят существенно сократить путь от идей до их материализации в цикле *информация — энергия — вещество*.

Единицами капитализации информации и знаний в фонде являются *конструктивные электронные издания*, построенные по схеме:

Объект ⇔ *Внешняя вычислительная среда* ⇔ *Абстрактная вычислительная среда* (Слои: *информационный, инструментально-технологический, дидактический, научно-исследовательский*) ⇔ *Внутренняя вычислительная среда* ⇔ *Субъект*.

Кроме того, поскольку фонд ориентируется на междисциплинарные проблемы и методы их решения, то он будет способствовать формированию интуитивного и научного мировоззрения, направленного на устойчивое развитие системы *природа — общество — человек*.

Рассмотрим ряд проблем, тенденций и ИТ-решений построения фонда.

Предметная (или объектная) инкапсуляция. В последнее время в связи с развитием рыночной экономики в российской научной среде стала проявляться тенденция к инкапсуляции предметного знания, причем в негативной форме «*натурального хозяйства*». Это происходит из-за недостаточности финансирования исследований и разработок, а также из-за большой трудоемкости и наукоемкости результатов профессиональной деятельности. Все это вынуждает реальных носителей знаний защищать результаты своего труда. Дополнительным негативным фактором такого подхода является снижение междисциплинарных связей, то есть возникает барьер для интеграции и гармонизации. Выход из создавшегося положения может быть найден в построении *иерархии технологий*, в создании *виртуальных коллективов*, а также в развитии меры и механизмов *оценки профессионального труда*.

Научное наследие. Потребность в глубоком понимании истоков, причин, среды появления и концентрации знаний, неравномерности или одновременности решений и научных открытий в разных географических точках и временных срезах, в управлении процессами развития и деградации. Кроме того, возникает потребность в построении барьеров для фальсификации или искажения знаний, в выработке более продуктивных критериев новизны и эффективных механизмов обнаружения ошибок и заблуждений (например, в опровержении трюизмов). Основой такой деятельности может стать научное наследие в форме электронных копий печатных изданий (редких, классических и ключевых). К сожалению, исторический опыт изобилует кострами и просто нелепыми решениями, когда книги уничтожались по году выпуска. Создание отсканированных копий печатных изданий сохраняет не просто форму (что тоже важно), но и создает почву для более детального и масштабируемого автоматизированного анализа («*технология обогащения руды*»). Кроме того, меняется и скорость доступа к

требуемой информации, когда распознанный текст совмещается с его графическим образом, например, продукт SST DjVu Toolkit.

О *формализации и доказательном построении*. Сейчас в публикациях преобладает «эмоциональный стиль» изложения. А эмоциональный — означает то, что структурированные и предметные части чаще всего тонут в дополнительных эмоциональных словах и выражениях, которые помогли автору выразить мысль, но которые часто мешают потребителю информации. В процессе формализации необходимо стремиться к созданию группы микроязыков, которые позволяют охватить развитие формы описания от возникновения идеи до ее реального воплощения. Такой опыт был получен при разработке системы «Ньютон» — группы согласованных языков: $tx - mi - module(x)$. Для инженерно-технических и естественно-научных электронных публикаций важно выделение как логической доказательной структуры текста, так и математической формы, включающей систему обозначений, формулы, определения, формулировки теорем и лемм, прагматические суждения, неформальные комментарии, прецеденты использования, демонстрационные конструкции, формирующие структуру математического доказательства (Г. С. Цейтин, Н. Н. Непейвода). Интерпретация перечисленных форм в рамках группы микроязыков в сочетании с методами логического вывода, а также интерактивно-динамическими представлениями создаст новую конативную, когнитивную, креативную, конструктивную и коммуникативную среду для автора и читателя. В рамках этой среды будут формироваться различные *стили объективизации* научного знания: личностный, корпоративный и парадигмальный (Ю. В. Чебраков).

О *гармоничном и уникальном развитии, о тьюнинге*. Решение сложных проблем всегда сопровождается максимальной концентрацией внимания и вовлечения всего интеллектуального и физического потенциала личности, подобно настройке музыкального инструмента (*тьюнинг*). Чтобы достичь такого состояния, придется долго настраиваться на проблему, вовлекая и удерживая ту информацию и знания, которые в конечном итоге приводят к озарению. Это состояние приходится проходить каждому, и помогает в этом гармоничное и уникальное развитие, дающее разносторонние ассоциации и точки зрения, а не просто наибольший финансовый ресурс так называемого успешного специалиста.

О *синтезе системного, либернетического и синергетического подходов*. Используемый системный подход предполагает построение пространства анализа и синтеза систем со следующими осями представления: *формы описания* (морфологическое, атрибутивное, функциональное, генетическое, целеполагание и механизмы построения); *уровни описания* (мировоззрение, теория, технология, практика); *проблемная ситуация* (субъект, объект, инструменты, цель, процессы, ресурсы, оценки). Чтобы более понятно и точно описать проблемную ситуацию, используется модифицированная система вопросов формирования элементов знаний (П. Г. Кузнецов, О. Л. Кузнецов, Б. Е. Большаков — см. таблицу).

№	Форма	Логика	Содержание	№
1	Зачем?	↔	Цель	1
2	Почему?	↔	Причина	2
3	Кто?	↔	Субъект	3
4	Что?	↔	Объект	4
5	Какой?	↔	Свойства	5
6	Как?	↔	Правила	6
7	Чем?	↔	Инструмент	7
8	Сколько?	↔	Ценность (Цена ресурса)	8
9	Где?	↔	Пространство	9
10	Когда?	↔	Время	10
№	Вопрос	Творчество	Ответ	№

Поиск на основе элементов знаний и системных представлений, в отличие от ключевого поиска, приводит к *смысловой индексации*, которая существенно сокращает и структурирует пространство поиска. Также сокращаются расстояния между идеей и ее материализацией, ликвидируются посредники между носителем знаний и их потребителями. В синтетическом подходе либернетика является инструментом нахождения новых степеней свобод и служит средством сокращения альтернатив, а синергетика является основой кристаллизации знаний и определения критических точек роста, в которых происходит переключение системы из режима параметрического управления в структурный режим, а затем и в базисный (принципиальный).

Список литературы

- Ильиных 2004 — *Ильиных, М. С.* Создание электронных изданий на основе системных представлений для поддержки профессиональной деятельности : автореф. дис. ... канд. техн. наук / М. С. Ильиных : ГОУ ВПО ИжГТУ. — Ижевск, 2004. — 20 с.
- Маслов 2001 — *Маслов С. Г.* Методология создания электронных учебных изданий / С. Г. Маслов, С. В. Моченов // Материалы XXIII науч.-метод. конф. ИжГТУ (Ижевск, 19–22 февр. 2001 г.). — Ижевск : Изд-во ИжГТУ, 2001 — С. 29–31.
- Maslov et al. 2005 — *Maslov, S. G.* Constructive Electronic Publication / S. G. Maslov, M. S. Ilinykh // Proceedings of the 7-th International Workshop on Computer Science and Information Technologies CSIT'2005. — Ufa, 2005 — P. 190–194.
- Маслов и др. 2005 — *Маслов, С. Г.* Среда конструктивной деятельности / С. Г. Маслов, Д. А. Дунаев, М. С. Ильиных // Технологии информатизации профессиональной деятельности : сб. труд. конф. — Ижевск : НИЦ «Регулярная и хаотическая динамика», 2005 — С. 155–195.

Glagolitic Alphabet in *Unicode*: Proposals for the Perfection of Slavonic Ranges

Heinz Miklas
University of Vienna, Austria

Несмотря на то, что существующая версия Unicode (4.1) представляет собой несколько более совершенную версию по сравнению с предыдущими в отношении передачи глаголического алфавита, она не является ни полной, ни достаточной в каком-либо ином отношении для использования. В частности, недостатки касаются (а) структуры диапазона, (b) инвентаря и расположения символов, (c) формы глифов, (d) имен букв и (e) числовых значений знаков.

Автор делает краткий обзор ранней истории глаголической системы, а затем выносит на обсуждение возможные способы решения существующих проблем.

Preliminaries

The present Unicode-version 4.1 of the Glagolitic writing system is neither complete nor in any other respect sufficient for its intended usage as a basic standard. In order to show this, the paper will shortly characterize the system and its early history and then propose a new version, discussing various options for the solution of special problems.

Theoretical requirements: A basic standard system must meet the following conditions:

- completeness,
- adequacy,
- practical functionality.

Since adequacy can only be achieved on a single historical level, the ordering, form and function of the units as well as their characterization (names of characters, etc.) have to depict the classic status of the Glagolitic system, reached during the Bulgarian development of O(ld) C(hurch) S(lavonic). Room provided, further elements (glyphs and/or functions) can be added to this set for reasons of practical functionality. They may be derived from the following phases of Glagolitic history (prehistoric phases marked by asterix *):