

Тематика информационных систем

в государственном образовательном стандарте
по направлению
«Информационные технологии»

Когаловский Михаил Рувимович

Институт проблем рынка РАН
МГУ, экономический факультет

kogalov@cemi.rssi.ru

Деятельность по стандартизации ИТ-образования

- Актуальность стандартизации содержания учебных программ в области ИТ-образования
- Результаты деятельности ACM и IEEE CS - Computing Curricula 2001 (CC-2001), проект CC-2005
- Текущая версия проекта Computing Curricula 2005 включает теперь уже 5 томов:
 - + Computer Science, 2001
 - + Computer Engineering, 2004
 - + Software Engineering 2004
 - + Information Systems, 2002
 - + Information Technologies, 2005
- Отечественный экспериментальный стандарт "Информационные технологии" – 511900 (2003 год)
- Слабые места стандарта, в частности, связанные с тематикой ИС
- Цель доклада.

ИС и информационные технологии (1)

- ИС - особый класс компьютерных систем
- В состав огромного многообразия информационных технологий входят пласты, специфичные для ИС - это технологии управления информационными ресурсами
- Сформировалось три магистральных направления таких технологий:
 - + технологии баз данных,
 - + технологии ИПС (комментарий)
 - + веб-технологии
- Каждое из этих направлений имеет дело со свойственной ему категорией информационных моделей и с особой категорией информационных ресурсов (ИР).

ИС и информационные технологии (2)

- Технологии управления информационными ресурсами используются не только в "чистом" виде, но и интегрируются при создании "гибридных" систем
- Сегодня имеются многие примеры БД, содержащих неоднородные ИР (таблицы, объекты, тексты, изображения, XML-документы, ...)
- Некоторые СУБД сегодня – многофункциональный "комбайн" для управления такими неоднородными ИР
- Предмет изучения в этой области включает теорию, на которой основаны те или иные технологии, существо этих технологий и их инструментальное воплощение, а также приложения
- ИС - актуальная область компьютерного дела
- Тематика ИС заслуживает полноценного представления в образовательных стандартах.

Тематика ИС в стандартах

- К сожалению, в СС-2001 и в отечественном стандарте по направлению "Информационные технологии" эта цель, на наш взгляд, пока не достигнута:
 - + отсутствует единый подход к обсуждению ИС
 - + не представлена целостная картина состояния и перспективных направлений развития технологий управления информационными ресурсами в ИС различных видов
 - + упущен ряд важных аспектов этих технологий
 - + имеются некорректности в содержании ряда специализаций и разделов некоторых дисциплин
 - + не рассматриваются экономические аспекты разработки ИС и управление такими системами
 - + имеются терминологические погрешности.

Почему нужен единый подход?

- В развитии технологий ИС в последнее десятилетие сложились две важные тенденции:
 - + интеграция информационных технологий в разработках информационных систем
 - + конвергенция технологий управления информационными ресурсами
- В связи с этим необходимо совместно изучать различные пласты технологий ИС
- Это невозможно без единого подхода к их изучению.

Интеграция технологий (1)

- Примеры интеграции технологий:
 - + доступ к системам баз данных в среде Веб
(Дж. Грей - веб-сервисы)
 - + XML-ориентированные базы данных
 - + реализация функций текстового поиска в СУБД
 - + текстовый поиск в Веб
- Еще пример: серверы баз данных ведущих поставщиков программного обеспечения систем баз данных уже не являются СУБД в классическом смысле
- Это функционально более сложные системы, в которых:
 - + управление структурированными данными - от СУБД
 - + новый вид данных, текстовый поиск, интеллектуализация интерфейсов - от технологий текстового поиска
 - + коммуникации, новый вид данных – от веб-технологий.

Интеграция технологий (2)

- Интеграция технологий связана не только со стремлением концентрировать разноаспектную функциональность в едином инструменте, но и с другими причинами:
 - + создание систем интеграции информационных ресурсов из множества неоднородных источников
 - + обеспечение интероперабельности и интеграции приложений, основанных на различных технологиях.

Конвергенция технологий

- Вторая важная тенденция - проявляющаяся в последние годы конвергенция различных пластов технологий ИС
- В частности, можно наблюдать процессы взаимопроникновения идей технологий баз данных и текстового поиска, веб-технологий и технологий баз данных, технологий текстового поиска и Веб (инвертированные списки в системах текстового поиска, идеи из БД в стандартах XML...)
- В этих условиях только при изучении базовых технологий с единых позиций можно полноценно изучить их существо и перспективы развития.

Единый подход (1)

- Основа единого подхода - представление об ИС как о действующих информационных моделях реальности (предметной области)
- На основе единого подхода у обучающихся необходимо сформировать целостное представление о всем возможном спектре ИС, об их разнообразии
- Это должно быть **концептуальное обсуждение**, независимое от конкретных инструментальных программных продуктов
- Для решения этой задачи лучше всего могла бы служить дисциплина "Information Management", которая включена в СС-2001, но охватывающая более широкий круг аспектов управления ИР и построенная на иной основе.

Единый подход (2)

- На наш взгляд эта дисциплина должна быть включена в учебные программы как обзорный курс, охватывающий в агрегированном представлении всю сферу технологий управления информацией
- Курс "Управление информацией" должен быть, на наш взгляд, основой для более детального обсуждения отдельных технологий этого назначения
- Такого рода курсы разной степени сложности читаются нами бакалаврам и магистрам на экономическом факультете МГУ с конца 90-х годов и постоянно актуализируется.

Примерное содержание обзорного курса (1)

- Краткая история развития технологий ИС
- ИС как информационная модель предметной области
- Инфологическая и даталогическая сфера ИС
- Классификация и свойства информационных моделей предметной области
- Ресурсы информационных систем
- Свойства информационных ресурсов (данных и метаданных) как "строительного материала" информационных моделей реальности
- Модели данных - инструмент моделирования предметной области и данных
- Задачи администрирования информационными ресурсами
- Архитектурные основы ИС, технологическая архитектура, информационная архитектура и архитектура данных.

Примерное содержание обзорного курса (2)

- Общая характеристика базовых технологий управления информационными ресурсами в ИС и сфер их применения
- Основы технологий баз данных, перспективные направления их развития
- Технологии ИПС, их эволюция, текстовые системы, системы текстового поиска, направления их развития
- Веб как глобальная гипермедийная распределенная ИС и коммуникационная среда, технологии Веб нового поколения, предпосылки их создания, направления развития
- Конвергенция технологий управления данными
- Интеграция данных из многих автономных независимых источников
- Технологии информационной безопасности в ИС
- Деятельность по стандартизации технологий управления данными.

Потерянные темы в 511900 (1)

- Не уделяется должного внимания обсуждению свойств информационных ресурсов, забыты метаданные как важная часть ИР, не рассмотрены их разновидности и стандарты
- Из стандартов управления ИР упоминаются главным образом лишь некоторые стандарты технологий БД
- Веб вообще не показан как информационная система. Из двух его функций уделяется внимание лишь одной – коммуникационной **в части разработки веб-приложений**
- Базовые аспекты Веб как уникальной глобальной гипермедийной распределенной системы не охвачены, не уделено должное внимание современным технологиям поиска в Веб, новой технологической платформе XML, XML-ориентированным базам данных, семантическому Веб, технологиям веб-сервисов
- Ограниченно рассматривается злободневная проблема интеграции неоднородных информационных ресурсов из множества автономных независимых источников.

Потерянные темы в 511900 (2)

- Не прослеживается эволюция ранних ИПС к современному их состоянию, не показаны различные классы таких систем и новые их функциональные возможности
- Не рассматриваются перспективные направления развития ИПС, направление Text Mining, не обсуждаются подходы к оценке качества текстового поиска
- Не обсуждаются важные классы систем баз данных - мобильные, статистические, XML-ориентированные, а также системы управления потоками данных
- Не рассмотрены и даже не упоминаются многомерные модели данных
- Не упоминаются также стандарты ODMG и ряд других стандартов де-факто в технологиях баз данных
- Не рассматриваются проблемы эталонного тестирования СУБД, стандарты ТРС (Transaction Processing Performance Council)
- Не дается оценки состояния технологий БД и важнейших перспективных направлений их развития.

Потерянные темы в 511900 (3)

- Упущены такие важные разделы технологий баз данных, как хранилища данных, технологии OLAP, Data Mining. Они не представлены, в частности, и в специализации "Управленческие информационные системы" (511915), где они играют ключевую роль
- Из "вертикальных" сфер применения информационных технологий представлена лишь биоинформатика
- Полностью отсутствует обсуждение особенностей геоданных, стандартов представления таких данных (в БД и в Веб), а также технологий ГИС
- Нет речи о грид-технологиях и, в частности, о гридах данных
- Ущербно выглядит специализация "Электронные библиотеки" (511907), в ней не рассматриваются ключевые элементы технологий ИС этого класса
- Не представлены вопросы экономики ИС и управления такими системами.

Примеры некорректностей (1)

- При заимствовании из СС-2001 раздела "Information Management" он превратился в следующее:

ОПД.Ф.08. Технологии баз данных (БД):

БД1: Информационные системы

БД2: Системы управления базами данных

БД3: Модели данных

- Собственно о базах данных и системах баз данных речь при этом не идет
- Дисциплина "Открытые информационные системы" (511911) - конгломерат самых разных аспектов информационных технологий: открытые системы вообще, некоторые фрагменты технологий ИС, сетевые технологии, языки и технологии программирования, архитектура CORBA и ее использование для решения проблемы унаследованных систем ...
- ИС здесь фактически подменяются компьютерными системами вообще.

Примеры некорректностей (2)

- Неадекватно названию и тематическое наполнение специализации "Управленческие информационные системы" (511915). Целый ряд рассматриваемых здесь вопросов не специфичен для этого класса ИС
- Вместе с тем, не рассмотрены хранилища данных, многомерные модели данных, разновидности технологий OLAP, технологии Data Mining и Knowledge Discovery from Databases (KDD)
- Языки SGML, HTML и XML предназначены вовсе не для описания контента информационных ресурсов "в сети Интернет" (специализация "Web-технологии" - 511924)
- Рассмотрение здесь среды Веб в большой степени подменяется рассмотрением Интернет
- Веб-технологии отождествляются с "Web servise design" (511924)
- Не адекватно содержанию название раздела "Моделирование информационных процессов" (ЕН.Ф.02.02).

Терминология

- В спецификации стандарта 511900 имеется ряд терминологических дефектов
- О русскоязычном эквиваленте термина **Computing**
- Трактовка **Computing** как **Информационные технологии** не адекватна содержанию рассматриваемого стандарта
- По нашему мнению:
$$\text{Computing} = \text{Computer Science} + \text{Information Technologies} + \text{IT Applications}$$
- Названия стандарта
"Компьютерное дело" или
"Компьютика"
точнее отражали бы его содержание (по аналогии с "Информатикой", "Кибернетикой", "Генетикой"...)

Имеющийся опыт

- На экономическом факультете МГУ с конца 90-х годов читается обзорный курс по технологиям ИС, соответствующий рассмотренным подходам
- Кафедрой экономической информатики факультета подготовлен комплекс учебных курсов по экономике информационных систем
- Эти курсы в течение ряда лет читаются в бакалавриате и магистратуре экономического факультета МГУ
- Выпущен комплекс учебников.

Заключение

- Создание образовательного стандарта по направлению "Информационные технологии", безусловно, будет способствовать совершенствованию подготовки высококвалифицированных специалистов данного профиля в отечественных вузах
- Имеющаяся экспериментальная версия стандарта нуждается в постоянной поддержке, обеспечивающей актуализацию его спецификаций в соответствии с новыми достижениями в компьютерном деле
- Важный вопрос – о статусе стандарта
- Хотя существует СС-2001, в Стенфорде читаются курсы в области ИС, тематика которых выходит за рамки стандарта (Дж. Ульмана, Г. Гарсиа-Молина, Дж. Видом и др.).

Литература

- *Когаловский М.Р.* Перспективные технологии информационных систем. - М.: Компания АйТи, ДМК Пресс, 2003. – 288 с.
- *Когаловский М.Р.* Энциклопедия технологий баз данных. - М.: Финансы и статистика, 2002. – 800 с.
- *Лугачев М.И., Анно Е.И., Когаловский М.Р., Липунцов Ю.П., Скрипкин К.Г., Смирнов С.Н., Смирнова Е.Е.* Экономическая информатика: Введение в экономический анализ информационных систем. - М.: Инфра-М, 2005. - 956 с.