

# Представления и преобразования гипертекстовой информации в специализированных расширяемых форматах

«Региональная информатика-2000»

М.Ю.Колодин  
(СПИИРАН)

# Об авторе



- М.Ю.Колодин:
- Научный сотрудник  
ИГИТО СПИИРАН
- Основные интересы:  
мета-системы, XML, TeX
- Адрес:  
[myke@iias.spb.su](mailto:myke@iias.spb.su) Адрес:  
[myke@iias.spb.su](mailto:myke@iias.spb.su),  
<http://myke.da.ru>

М.Ю.Колодин.

# Тема и метод



- Область: обработка сложной информации в текстовой форме
- Особенности: сложноструктурированный гипертекстовый характер
- Методы: внешняя и внутренняя обработка представлений на основе XML

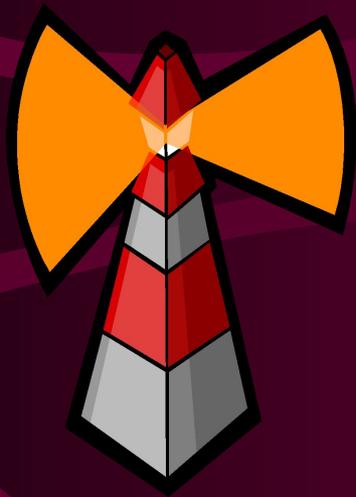
М.Ю.Колодин.

РИ-2000.

05.12.2000

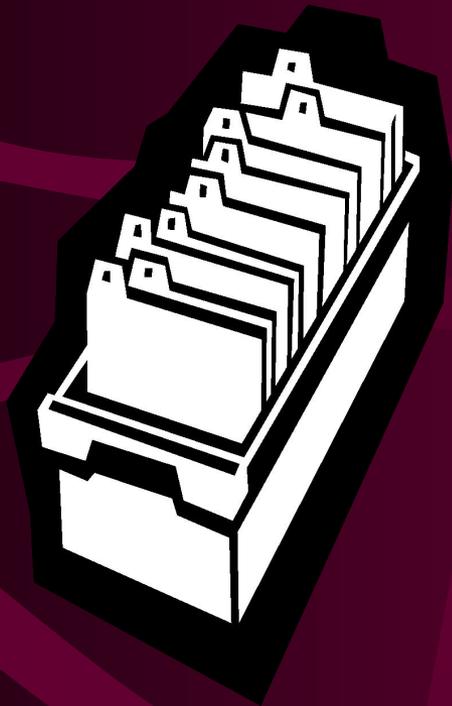
3

# Два типа языков



- Есть два способа представления информации в языковой форме:
- а. конкретные языки (конечные, специализированные)
- б. мета-языки (с порождением производных языков)

# Простые языки



- Частные языки
- Новые программы обработки для каждого языка
- Доказательства (обоснования) для каждого языка
- Оптимизация
- Ограниченность применения
- Широкое использование

М.Ю.Колодин.

РИ-2000.

05.12.2000

5

# Мета-языки



- Описания семейства языков
- Общие обработчики
- Единая доказательная база
- Многократность использования
- Многообразие представлений
- Адекватность предметной области

М.Ю.Колодин.  
РИ-2000.

05.12.2000

6

# Примеры мета-языков

- SGML, основной мета-язык
- SGML-производные (DocBook, TEI, пр.)
- XML, новый мета-язык
- XML-производные (MathML, XLink, XPointer, XSL, XSLT,..)



# Традиционные обработчики



- Модели: SAX, DOM
- Языки: C++, Java, Perl, Python, прочие
- Внешний просмотр документа с преобразованием его в HTML, XML, пр.

М.Ю.Колодин.  
РИ-2000.

# Новые методы



- Внутренние (непосредственные) обработчики
- Настройка инструментального средства (FORTH)
- Многоуровневые представления

# Инструментальные средства



- FORTH: переопределение синтаксиса языка для включения целевого языка
- Непосредственное исполнение текста на целевом языке в различных контекстах
- Получение различных производных

М.Ю.Колодин.

РИ-2000.

05.12.2000

10

# Многоуровневые представления



- Мета: «между», «над»
- Описание систем
- Многоуровневое (произвольно-уровневое) описание
- Обобщение системного представления

М.Ю.Колодин.  
РИ-2000.

# Результаты



- Значительная польза для ТЕОРИИ (единый аппарат, формальная модель, систематизация, прочее)
- Ощутимая ПРАКТИЧЕСКАЯ польза (эффективность, применимость, доступность, прочее)

М.Ю.Колодин.

РИ-2000.

05.12.2000

12

# Новые исследования



- Сравнение с другими обработчиками (Java, C++, Perl, Python, пр.)
- Обработка систем сложноструктурированных текстов с ГТ-разметкой
- Интернет-приложения
- Развитие теории

М.Ю.Колодин.

РИ-2000.

05.12.2000

13

# Ссылки



- Языки ГТ-разметки (SGML, XML, пр.): <http://www.w3c.org>;
- Обработчики: Java (<http://java.sun.com>), Perl (<http://www.perl.com>);
- Мои материалы: <http://myke.webjump.com>

М.Ю.Колодин.

РИ-2000.

05.12.2000

14

Спасибо!



М.Ю.Колодин.  
РИ-2000.

05.12.2000

15