

Содержание

ВВЕДЕНИЕ	12
ГЛАВА 1. NEW MEDIA	15
1.1. Общая информация об Internet	15
1.2. New Media и СМИ	17
1.3. Гипертекст и WWW	19
1.4. Интеграция информационных ресурсов	20
1.5. Топология Web-пространства	23
1.6. Навигация в Internet	25
1.7. Информационно-поисковые системы	28
1.8. “Скрытый” Web	31
1.8.1. Очередной феномен Internet	31
1.8.2. Типы скрытых ресурсов	33
1.8.3. Базы данных “скрытой” Сети	34
1.8.4. Сталкеры в скрытом пространстве	37
1.8.5. “Скрытый” Web в каталогах	38
1.8.6. Системы поиска в “скрытом” Web	39
1.8.7. Информация в различных форматах	40
1.8.8. Скрытые новостные ресурсы	40
1.8.9. “Скрытый” архив “поверхностного” Web	41
1.8.10. Подходы к решению проблемы “скрытого” Web	41
ГЛАВА 2. ПОИСК В INTERNET	43
2.1. Характеристики ИПС	43
2.2. Лингвистическое обеспечение ИПС	45
2.3. Семантические методы	49
2.4. Этапы поисковой процедуры	52
2.5. Процесс поиска непосредственно	54
2.6. Запросы пользователей	55
2.7. Поиск подобных документов	57
2.8. Ранжирование откликов	57
2.9. Поиск по словам и словоформам	57
2.10. Логические операторы	58
2.11. Операторы контекстной близости	59

2.12. Поиск по параметрам	59
2.13. Популярные сетевые информационно-поисковые службы	61
2.13.1. Крупнейшие зарубежные службы	61
2.13.2. Службы поиска в российском сегменте Сети	68
2.13.3. Крупнейшие украинские службы	70
2.14. Поиск информации в корпоративных сетях	73
2.14.1. Популярные ИПС	73
2.14.2. Новый уровень обработки сетевой информации	79
2.14.3. Порталы знаний	81
2.15. Поисковые программно-аппаратные комплексы	83
ГЛАВА 3. СИСТЕМЫ ИНТЕГРАЦИИ INTERNET-КОНТЕНТА	87
3.1. Статическая и динамическая составляющие Web-пространства	87
3.2. Недостатки традиционного поиска	88
3.3. Невизуальный Web	89
3.4. Синдикация новостной информации	91
3.5. От “поисковиков” — к “интеграторам”	91
3.6. Форматы синдикации новостей	93
3.7. OPML — формат для хранения списка RSS-фидов	96
3.8. Источники новостного контента	98
3.9. Системы поиска RSS-фидов	104
3.10. Агрегаторы	106
3.11. Новые подходы	109
3.12. Информационные ресурсы для мобильных устройств	110
3.12.1. Wireless Application Protocol	110
3.12.2. WAP-ресурсы	111
3.12.3. Реализация WAP-протокола	113
3.12.4. WML и микробраузеры	114
3.12.5. Эмуляторы WAP	116
3.12.6. Проблемы и перспективы WAP	118
3.12.7. Доступ к сетевому контенту с КПК	121
3.12.8. Информационные ресурсы для КПК	122
3.12.9. Эмуляция мобильности	124
3.12.10. RSS-формат на КПК	125
3.12.11. Игрушка или рабочий инструмент	126
3.13. Службы доставки новостей по электронной почте	127
3.13.1. История сервиса	127
3.13.2. Система телеконференций Usenet	128
3.13.3. Доставка новостей с отдельных сайтов	131
3.13.4. Специализированные службы рассылки новостей	133
3.13.5. Интеграция новостей с целью рассылки	135

3.13.6. Спам — альтернатива востребованной рассылке	139
3.13.7. Перспективы технологий доставки новостей	139
ГЛАВА 4. XML — ЯЗЫК РАЗМЕТКИ И МОДЕЛЬ ДАННЫХ	141
4.1. XML как модель данных	144
4.2. XML-поиск и языки запросов	145
4.3. XML-решения для хранения данных	149
4.4. Корпоративные и офисные приложения для XML	154
4.5. Настоящее и обозримое будущее XML	156
ГЛАВА 5. ОСНОВЫ ТЕХНОЛОГИИ TEXT MINING	159
5.1. Основные элементы Text Mining	161
5.2. Контент-анализ	162
5.3. Модели поиска	166
5.3.1. Булева модель поиска	166
5.3.2. Векторно-пространственная модель	168
5.3.3. Гибридные модели поиска	169
5.4. Группировка текстовых данных	169
5.4.1. Кластеризация	171
5.4.2. Тематическая близость	172
5.4.3. Вероятностная модель	174
5.4.4. Латентно-семантический анализ	178
5.5. Автоматические ответы на вопросы	188
5.6. Реализация систем Text Mining	190
5.6.1. Intelligent Miner for Text	191
5.6.2. PolyAnalyst	192
5.6.3. Text Miner	194
5.6.4. SemioMap	195
5.6.5. InterMedia Text, Oracle Text	196
5.6.6. Autonomy IDOL Server	196
5.6.7. Galaktika-ZOOM	197
5.6.8. InfoStream	198
5.7. Text Mining не только для спецслужб	198
5.8. Автоматическое реферирование	199
5.8.1. Квазиреферирование	201
5.8.2. Алгоритмы автореферирования	202
5.8.3. Дайджесты	203
5.8.4. Поисковые образы документов	205
5.8.5. Информационные портреты	205
5.8.6. Программы автореферирования	205
5.8.7. Автореферирование на основе семантических методов	212
5.8.8. Перспективы автореферирования	214

ГЛАВА 6. ИНСТРУМЕНТАРИЙ КОНКУРЕНТНОЙ РАЗВЕДКИ	217
6.1. Задачи конкурентной разведки	218
6.2. Источники информации и базы данных	219
6.3. Подходы к анализу контента	220
6.4. Некоторые примеры	221
6.5. Конкурентная разведка и “скрытый” Web	227
6.6. Перспективы систем конкурентной разведки	227
ГЛАВА 7. ЗАКОНОМЕРНОСТИ, ПРИСУЩИЕ ИНФОРМАЦИОННЫМ СИСТЕМАМ	231
7.1. Правило Парето	231
7.2. О переходе количества в качество	233
7.3. Закон Зипфа	234
7.4. Закономерность Брэдфорда	238
7.5. Прогноз Мура и информационная сфера	239
7.6. Фракталы и информационное пространство	240
7.6.1. Примеры абстрактных фракталов	241
7.6.2. Фракталы из жизни	244
7.6.3. Информационные фракталы	245
7.7. Проблемы и феномены Internet	249
ГЛОССАРИЙ	253
ЛИТЕРАТУРА	263
ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ	267