

ТЕХНОЛОГІЇ XML У ПАТЕНТНО-КОН'ЮНКТУРНИХ ДОСЛІДЖЕННЯХ

Ю.А. Альошкіна, О.С. Дерев'янку

НТУ „ХПР”, м. Харків

В [1] запропоновано архітектуру системи автоматизації патентно-кон'юнктурних досліджень (САПКД), яка будується як ланцюжок модулів інтелектуального оброблення даних із зворотними зв'язками. Основною концепцією вибору засобів реалізації САПКД декларується базування їх на відкритих стандартах.

Інформація, яка оброблюється в САПКД, є як структурованою, так і напівструктурованою. Загальноприйнятим засобом представлення останньої є мова XML. Саме технології XML виявляються придатними для реалізації майже усіх ланок ланцюжка оброблення даних в САПКД.

Для «шлюзів» між джерелами патентної інформації та сховищем даних САПКД може застосовуватися технологія перетворення XSLT. Сховище даних САПКД має поряд з реляційними можливостями бути орієнтоване на зберігання також XML-документів і підтримувати мову запитів XQuery та суміщення мов XQuery та SQL відповідно до стандарту SQL/XML. Сама інфраструктура ланцюжка модулів реалізована у проекті UIMA, де визначено також XML-структуру анотації (CAS). Для представлення семантики анотацій пропонуються XML-базовані специфікації засобів представлення знань у «семантичному Web»: RDF та OWL. Для візуалізації результатів може застосовуватися мова HTML, засобом перетворення представлень результатів можуть бути шаблони XSLT. При реалізації САПКД у корпоративному середовищі інфраструктура взаємодії розподілених частин може бути забезпечена через Web-сервіси, у складі бизнес-процесу, описаного засобами BPMN і/або BPEL.

Наш вибір СУБД зупиняється на IBM DB2, яка суміщає реляційне сховище даних із сховищем даних XML у їх «природній» моделі, повний обсяг підтримки XQuery та SQL/XML та підтримку Web-сервісів. Для програмних компонентів системи пропонується платформа Java та інструментальна платформа Eclipse. В останній маємо підключення для майже усіх згаданих XML-технологій. Продукт IBM DB2 Data Studio також оснований на Eclipse і забезпечує доступ до технологій DB2 та встановлення нових підключень. Саме цей продукт може служити єдиною точкою доступу до усіх технологій в САПКД.

Література: 1. Дерев'янку А.С., Сомхиева О.С. Применение средств интеллектуальной обработки данных в патентно-конъюнктурных исследованиях. – Харків: Харківський університет Повітряних Сил, 2009, с. 162-167.