

Виконання плану заходів щодо її реалізації сприятиме подальшому розвитку та удосконаленню інформаційних відносин у суспільстві, зокрема побудові єдиної державної системи правової інформації, в якій електронні фонди нормативно-правової інформації парламенту, президента, уряду, міністерств та інших центральних органів виконавчої влади будуть поєднані між собою єдиними технологіями ведення, представлення, опрацювання, зберігання, актуалізації та оприлюднення нормативно-правової інформації на офіційних веб-сайтах цих органів влади, а впровадження та використання електронного цифрового підпису надасть юридичну силу електронним документам.

Андрій Баранцев

ПОШУК ТА ВІДОБРАЖЕННЯ ІНФОРМАЦІЇ НА РАСТРОВИХ ЗОБРАЖЕННЯХ ТЕКСТІВ, ЩО МІСТЯТЬСЯ В ЕЛЕКТРОННИХ ДОКУМЕНТАХ

Одним із поширених способів створення електронного документа є переведення документа з паперовим носієм в електронний вигляд і призначення йому відповідних обов'язкових атрибутів. Результатом такого переведення буде електронний документ, що містить растрове зображення, яке містить текст. При роботі з таким документом було б зручно мати такий же набір можливостей по повнотекстовому пошуку і наочному відображенню його результатів, як і при роботі із звичайним текстовим документом.

Провідні світові проекти по створенню та накопиченню електронних ресурсів, (Google Book, проект Національної бібліотеки Франції Gallica), надають користувачам такі можливості для растрових зображень. Це свідчить про те, що в них є потреба. Проте характерною особливістю цих зображень є необхідність використання складних процедур для пошуку і виділення необхідної інформації, а проекти, які успішно реалізували такі процедури, не надають опису їхньої реалізації. Нами описується одна з можливих реалізацій таких процедур.

Для реалізації описаних вище можливостей пропонуємо попередньо розпізнати текст на растровому зображенні таким чином, щоб визначити не тільки букви, але й додаткові дані, зокрема координати букв на зображенні. Нами визначено перелік обов'язкових (для виконання поставленого завдання) атрибутів, які підлягають формалізації після розпізнання тексту. Ці атрибути необхідно зберегти у файл фор-

мату XML певної структури. Такий файл названо текстовою мапою, що містить усі дані, необхідні для повнотекстового пошуку слова або його частини в електронному документі та визначення координат на зображенні, по яких розміщується знайдене слово. Для наочного відображення результатів пошуку необхідно виокремити текст на растровому зображенні. Для цього треба зробити копію вихідного зображення і відредагувати його, використовуючи координати знайдених слів із текстової мапи. Редагування зображення зводиться до заливання регіонів, що відповідають знайденим словам, визначеним кольором. Для збереження оригінальних зображень самих символів слід відрізнити символ від фону і залити певним кольором тільки фон. Відрізнити символ від фону можна за контрастом. Останній визначається відношенням яскравості символу і фону. На зображеннях текстових документів яскравість фону вище яскравості символу. Тому при редагуванні зображення необхідно вибрати певний поріг яскравості і залити кольором тільки ті точки растра, яскравість яких вище порога. Поріг яскравості встановлюється індивідуально для кожного документа і міститися в текстовій мапі.

Для практичного забезпечення можливості повнотекстового пошуку та наочного відображення його результатів автором розроблено неінтерактивний графічний редактор-маркер зображення E-Cologer, демонстраційна сторінка якого знаходиться на електронному ресурсі за адресою: <http://micrography.gov.ua/ecodemo>. Основну частину демонстраційної сторінки займає растрове зображення документа у форматі JPEGi, на якому демонструються можливості редактора E-Cologer. У лівій частині сторінки є поле для введення тексту, пошук і виділення якого буде проводитися після натискання кнопки «Применить». Також у лівій частині демонстраційної сторінки розташовані елементи управління, за допомогою яких демонструються додаткові можливості редактора E-Cologer. Доступ до редактора E-Cologer організовано через CGI-сценарій. При натисканні кнопки «Применить» редактор E-Cologer завантажує текстову мапу і будує по ній об'єктну модель растрового зображення. Повнотекстовий пошук та визначення координат регіонів для маркування тексту виконується по об'єктній моделі, а не по файлу текстової мапи. Такий підхід прискорює роботу редактора. Потім редактор E-Cologer творить копію вихідного растрового зображення і маркує обчислені регіони кольором, указаним як колір маркера. Відредаговане таким чином зображення редактор E-Cologer повертає CGI-сценарію.

Вочевидь, що для кожного растрового зображення треба створити індивідуальну текстову мапу. Ручна побудова їх – трудомісткий процес. А використання редактора E-Cologer може бути ефективним лише при наявності простого способу створення текстових мап. Тому нами у межах проекту E-Cologer розроблено редактор текстових мап E-Mapedit, який автоматизує процес їх створення. Редактор E-Mapedit розпізнає текст вихідного зображення і в автоматичному режимі будуватиме цифрову мапу. Результат роботи завантажуватиметься у вікно редагування, в якому користувач має можливість у ручному режимі відкоригувати текстову мапу та виправити помилки розпізнавання, які можуть мати місце при поганій читаності вихідного документа. Демонстраційна версія редактора E-Mapedit доступна для завантаження з електронного ресурса за адресою: <ftp://micrography.gov.ua/pub/demo/ecologer>. Для встановлення демонстраційної версії редактора на персональний комп'ютер необхідно скористатися інструкцією [howto-install.pdf](#), яка знаходиться за вказаною вище адресою.

Підхід до формалізації атрибутів розпізнаного тексту растрових зображень, запропонований нами, відкрив можливість реалізувати механізм повнотекстового пошуку і наочного відображення його результатів на растрових зображеннях текстів, що містяться в електронних документах. Основними перевагами реалізованого механізму є:

- простота у використанні;
- невимогливість до обчислювальних ресурсів операційної системи;
- використання відкритого формату XML.

Крім використання реалізованого механізму для поліпшення якості забезпечення документами користувачів електронних архівів, перспективними також є можливості його використання для створення електронних бібліотек і в більшості напрямків формування та використання страхового фонду документації України.