

УДК 004.91

ПРОБЛЕМА ПРОГРАМНОГО РЕДАГУВАННЯ ДОКУМЕНТІВ ФОРМАТУ MICROSOFT WORD

Р. О. Попов, студент

Дніпровський національний університет імені Олеся Гончара

Н. В. Карпенко, канд. фіз.-мат. наук, доцент

Дніпровський національний університет імені Олеся Гончара

Ключові слова: Microsoft Word, програмне редагування документів, шаблони документів, автоматизація документообігу, інформаційні технології.

Наразі існує велика проблема в сфері освіти України, яка напряду не зв'язана с викладанням – це створення та обробка документації, що супроводжує освітній процес. Хоча інформаційні технології спростили створення документів в електронному форматі, всеодно витрачається багато часу на заповнення документації. Складнощі виникають, коли потрібно створювати кілька документів для одного об'єкта з предметної області, наприклад, різні заяви, анкети та контракти для одного студента на конкурсну пропозицію під час вступу до закладу вищої освіти. Таким чином, виникає ідея розробки програмного забезпечення для автоматизації цих рутинних завдань із використанням спеціалізованих бібліотек для організації шаблонів документів та їх заповнення.

Оскільки де-факто стандартом для створення електронних документів в Україні є програма Microsoft Word, у цій роботі ми розглянемо наявні інструменти програмного редагування для створення таких документів та особливості створення програмного забезпечення для обробки таких файлів.

Microsoft Word – це текстовий процесор розроблений компанією Microsoft, який входить до пакету Microsoft Office. Microsoft Word зберігає документи у форматі .docx за специфікацією Office Open XML [1]. У табл. 1 наведено бібліотеки для обробки .docx файлів для різних мов програмування.

З огляду на ці бібліотеки, можна зробити висновки, що екосистема редагування документів не є розвинуеною та популярною в усіх мовах програмування. Бібліотеки з повним пакетом можливостей редагування документів розробляються комерційними компаніями та досить коштовні для користування. А розробники-початківці, які працюють з мовою C#, змушені користування безкоштовною версією. А це досить незручно, оскільки, наприклад, GemBox.Document обмежує безкоштовне користування 20 рядками тексту. Хоча інколи безкоштовні продукти не уступають комерційним, як в Python (Spire.Doc та python-docx). Існує також проблема, що бібліотеки в одній мові програмування можуть бути більш потужними, аніж в інших мовах.

Саме за цих причин іноді редагування та генерація документів в форматі .docx стає непередбачуваною проблемою. В одних ситуаціях відредаговане поле має неправильне форматування, в інших – документ розтягується до більшої кількості сторінок, ніж дозволено.

Таблиця 1 - Бібліотеки обробки .docx файлів

Мова програмування	Назва бібліотеки	Умови використання	Активність розробки
C#	GemBox.Document	Лімітована безкоштовна версія та повна платна	Активно розробляється компанією GemBox
	DocX	Лімітована безкоштовна версія та повна платна	Активно розробляється компанією Xseed
JavaScript	docx.js	Безкоштовно	Розробляється спільнотою
Python	python-docx	Безкоштовно	Останнім часом неактивно розроблюється спільнотою
	Spire.Doc	Платно	Розробляється компанією (має схожий функціонал з python-docx)
C++	DuckX	Безкоштовно	Останнім часом неактивно розробляється спільнотою

Лімітованість бібліотек також змушує переформатовувати документи в іншу форму, яка є зрозумілою для бібліотеки. Ще однією особливості розробки шаблонів у .docx є те, що системи контролю версій не можуть визначити конкретні зміни у файлі (рядки, що змінились), що дуже сповільнює процес розробки програмного забезпечення.

Причиною таких недоліків є те, що OOXML (поточний стандарт .docx) хоча і є відкритим, але є дуже складним (на відміну від вже існуючого Open Document) [2]. Для розробки повної бібліотеки необхідно дуже багато часу та зусиль, але фінансування для такої праці мають лише великі корпорації. Вже те, що Word, насамперед, налаштований на редагування в стилі WYSIWIG, робить програмне редагування дуже складним.

Список використаних джерел

1. International Organization for Standardization Information technology — Open Document Format for Office Applications (OpenDocument) v1.2. 2015. URL: <https://www.iso.org/standard/66363.html>
2. Starynkevitch B. Comment of Basile Starynkevitch on StackOverflow question. 2018. URL: <https://opensource.stackexchange.com/a/6481>