

УДК 681.3;377.4

Петрушко Василь Андрійович, провідний інженер відділу електронних інформаційних ресурсів і мережних технологій Інституту інформаційних технологій і засобів навчання НАПН України, м. Київ

РОЗРОБКА ЗАСОБІВ СИНХРОНІЗАЦІЇ ДАНИХ ДОКУМЕНТІВ НА ПЛАТФОРМІ SHAREPOINT

Анотація

Створення систем автоматизації документообігу для управління освітніми процесами — це актуальне завдання в умовах розвитку інформаційного суспільства. Однією із сучасних платформ для створення подібних систем є Microsoft SharePoint Products and Technologies. Sharepoint містить бібліотеки для зберігання документів, зручні інструменти для редагування і керування версіями документів. У системах документообігу виникає потреба повторного використання даних у різних документах. Sharepoint не має вбудованих засобів для синхронізації даних між документами. У статті описано розробку засобу для автоматичної синхронізації даних між документами у Sharepoint, на основі якого здійснюється розробка інформаційної системи «Планування наукових досліджень в Національній академії педагогічних наук України на базі мережі Інтернет».

Ключові слова: синхронізація даних, документообіг, SharePoint, інформаційна система, Інтернет, портал.

Вступ

Microsoft SharePoint Products and Technologies

Microsoft SharePoint Products and Technologies SharePoint (далі Sharepoint) — це система керування вмісту з інтегрованим функціоналом пошуку.

Sharepoint — це колекція програмних продуктів і компонентів, яка включає такі основні елементи:

- набір веб-застосунків для організації сумісної роботи;

- функціонал для створення порталів;
- модуль пошуку інформації в документах та інформаційних системах;
- функціонал керування робочими процесами та систему керування корпоративним вмістом;

- модуль створення форм для введення інформації;
- функціонал для бізнес-аналізу.

Платформу SharePoint можна використовувати для створення сайтів, що пропонують користувачам можливості для сумісної роботи. Створені на платформі Sharepoint сайти можна використовувати як сховища інформації, знань і документів, а також для виконання веб-застосунків взаємодії, таких як «вікі» та блоги. Користувач може керувати і взаємодіяти з інформацією у списках і бібліотеках, використовуючи елементи керування, які називаються веб-частини [1].

Платформа Sharepoint пропонує різноманітні можливості розширення функціоналу для розробників ПЗ.

Метою даної роботи є аналіз проблем, що виникають під час розробки коду для синхронізації даних між документами на платформі Sharepoint.

Інформаційна система «Планування наукових досліджень в Національній академії педагогічних наук України на базі мережі Інтернет»

Основним завданням інформаційної системи «Планування наукових досліджень в Національній академії педагогічних наук України на базі мережі Інтернет» (далі ІС «Планування») — створення та оновлення супровідних документів для науково-дослідних робіт (НДР). Нині в ІС «Планування» налічується близько 20 різних типів документів. Більшість даних (полів) цих документів повторюються. Наприклад: назва НДР, ПІБ керівника тощо. Для максимальної зручності в ІС «Планування» користувач вносить кожне унікальне поле один раз. На жаль, у платформі Sharepoint немає вбудованої підтримки синхронізації даних між документами. Для цього було розроблено власне програмне рішення. Деякі особливості Sharepoint не дозволяють зробити це тривіально. У цій статті розглядається специфіка написання коду синхронізації документів для MS Sharepoint, проблеми, які виникали під час розробки, і їх рішення.

Вибір формату документів

Процес синхронізації даних між документами потребує серверної обробки документів. Бінарний формат документів, який використовується в системі MS Office до версії MS Office 2007, закритий. Тому його редагування потребує запуску MS Office, що використовує дуже багато ресурсів. Але з появою формату MS Office 2007 (розширення DOCX) це завдання набагато спростилося. Формат документів MS Office 2007 – OOXML (Office Open XML), який базується на текстовому форматі XML. Формат XML відкритий і дуже популярний, тому його обробка реалізована майже в усіх мовах програмування. Також обробка текстового формату XML значно менш ресурсоемна, ніж обробка закритого бінарного формату DOC. Тому серверна обробка документів набагато ефективніша для документів DOCX. Також для зручнішого обміну даними у форматі DOCX додана підтримка експрес-блоків [2].

Експрес-блоки

Експрес-блоки (quick parts) — це текст для повторного використання, збережений у спеціальній колекції.

Під час використання платформи Sharepoint у поєднанні з програмою MS Word 2007 чи пізнішої версії починає працювати вбудована синхронізація експрес-блока «поле» (Quick parts -> Field) з типом вмісту документа. Експрес-блоки дуже зручні для серверної обробки даних документів, тому система ІС «Планування» підтримує тільки новий формат документів MS Office 2007 (файли з розширенням DOCX) [3].

Зберігання документів у Sharepoint для ІС «Планування»

Для зберігання документів Sharepoint пропонує бібліотеку документів. Бібліотека документів містить елементи типу вмісту «документ». Для системи ІС «Планування» були створені типи вмісту для всіх супровідних документів НДР за допомогою розширення типу вмісту «документ» необхідними полями [1].

Для зберігання даних НДР використовуються поля (field). Кожне поле зберігає окрему сутність НДР, наприклад: тема НДР; науковий напрям, проблема дослідження, завдання програми тощо. НДР складається з набору унікальних полів. Кожен супровідний документ включає у себе сукупність цих полів.

Кілька прикладів установлення полів для НДР в Sharepoint (уривки з файлу CustomFields.xml):

Тема НДР:

```
<Field ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4001}"  
  Type="Text"  
  Name="z1"  
  DisplayName="Тема НДР"  
  StaticName="TemaNDR"  
  Hidden="FALSE"  
  Required="FALSE"  
  Sealed="FALSE"  
  Group="Planning" />
```

Науковий ступінь наукового керівника:

```
<Field ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4011}"  
  Type="Choice"  
  Name="z8"  
  DisplayName="Науковий ступінь наукового керівника"  
  StaticName="KerivnykStupin"  
  Hidden="FALSE"  
  Required="FALSE"  
  Sealed="FALSE"  
  Group="Planning">
```

<CHOICES>

<CHOICE>д. пед. н.</CHOICE>

<CHOICE>д. політ. н.</CHOICE>

<CHOICE>д. соц. н.</CHOICE>

<CHOICE>д. т. н.</CHOICE>

<CHOICE>д. ф.-м. н.</CHOICE>

<CHOICE>д. філол. н.</CHOICE>

<CHOICE>д. філос. н.</CHOICE>

<CHOICE>д. психол. н.</CHOICE>

<CHOICE>к. політ. н.</CHOICE>

```

<CHOICE>к. психол. н.</CHOICE>
<CHOICE>к.т.н.</CHOICE>
<CHOICE>к. ф.-м. н.</CHOICE>
<CHOICE>к. філол. н.</CHOICE>
<CHOICE>д. філос. н.</CHOICE>
</CHOICES>
</Field>

```

Поля використовуються в типах вмісту для супровідних документів НДР.

Приклад: тип вмісту для документа «Запит» (файл ZContentType.xml):

```

<?xml version="1.0" encoding="utf-8"?>
<Elements Id="85c8113b-1af8-4374-adf8-c5b62ff0184d"
xmlns="http://schemas.microsoft.com/sharepoint/">
  <ContentType ID="0x0101005961a6c8a2534fd0b72dbb4ae2667006"
    Name="Запит"
    Group="Документи"
    Description=""
    Version="0">
    <FieldRefs>
      <!-- Назва наукової установи -->
      <FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4023}" />
      <!-- Реєстраційний номер запиту -->
      <FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4058}" />
      <!-- Дата реєстраційного запиту -->
      <FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4059}" />
      <!-- ПІБ особи, що зареєструвала запит -->
      <FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4060}" />
      <!-- Тема НДР -->
      <FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4001}" />
      <!-- Науковий напрям, проблема дослідження, завдання програми -->
      <FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4013}" />
    </FieldRefs>
  </ContentType>
</Elements>

```

<!-- Номер протоколу постанови Президії НАПН про оголошення конкурсу наукових проектів -->

<FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4068}" />

<!-- Дата протоколу постанови Президії НАПН про оголошення конкурсу наукових проектів -->

<FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4069}" />

<!-- Номер протоколу Постанови вченої ради про виконання НДР -->

<FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4004}" />

<!-- Дата Постанови вченої ради про виконання НДР -->

<FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4005}" />

<!-- Термін виконання початок -->

<FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4014}" />

<!-- Термін виконання кінець -->

<FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4015}" />

<!-- ПІБ наукового керівника -->

<FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4009}" />

<!-- Посада наукового керівника -->

<FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4010}" />

<!-- Науковий ступінь наукового керівника -->

<FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4011}" />

<!-- Вчене звання наукового керівника -->

<FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4012}" />

<!-- Інформація про керівника -->

<!-- Виконавці -->

<!-- Календарний план робіт за темою -->

<!-- Виконавець -->

<!-- Орієнтовна вартість робіт за темою, тис. грн. -->

<FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4018}" />

<!-- Орієнтовна вартість роботи за роками -->

<FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4019}" />

```

<!-- ПІБ директора -->
<FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4002}" />
<!-- ПІБ головного бухгалтера -->
<FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4042}" />
<!-- Назва теми наукового дослідження (рос.) -->
<FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4070}" />
<!-- Назва теми наукового дослідження (англ.) -->
<FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4071}" />
<!-- Мета роботи -->
<FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4072}" />
<!-- Науковий керівник теми (рос.) -->
<FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4073}" />
<!-- Науковий керівник теми (англ.) -->
<FieldRef ID="{4BC3FA7F-DFCF-4927-BCB5-4B5F430C4074}" />
</FieldRefs>
</ContentType>
</Elements>

```

Тобто кожен тип вмісту документа НДР — це набір посилань на поля.

Під час завантаження чи оновлення документа з відповідними експрес-полями в Sharepoint вони автоматично синхронізуються з полями типу вмісту.

Оскільки поля в документах використовуються повторно, а Sharepoint не пропонує засобів для подібної синхронізації, був розроблений код, який синхронізує значення полів (і автоматично експрес-блоків у документі) між типами вмісту.

Розробка синхронізації даних документів

Для оновлення документів був написаний обробник події ItemUpdated (він викликається після збереження елемента в бібліотеці) [3].

Перша версія коду мала такий вигляд:

```

public override void ItemUpdated(SPItemEventProperties properties) // Елемент
додано
{

```

```

        SPLListItem item = properties.ListItem; // Витягнення оновленого елемента
(документа)
        SPLListItemCollection listItems = properties.ListItem.ListItems.List.Items; //
Вибір усіх елементів (документів) поточної бібліотеки (документів)
        foreach (SPLListItem itemTmp in listItems) // Для всіх елементів (документів) в
поточній бібліотеці (документів)
        {
            if (itemTmp.ID != item.ID) //Крім себе
            {
                foreach (SPField field in item.Fields) //Для кожного поля
                {
                    SPField fieldTmp =
itemTmp.Fields.GetFieldByInternalName(field.InternalName); // Спроба знайти таке саме
поле
                    if ( (fieldTmp != null) && !fieldTmp.FromBaseType &&
!fieldTmp.ReadOnlyField) // Поле знайдено, не наслідуване, не тільки для читання
                    {
                        itemTmp[field.InternalName] = item[field.InternalName]; //
Оновлення поля
                    }
                }
            }
            itemTmp.SystemUpdate(false); // Збереження елемента (документа)
        }
    }
}

```

Під час тестування з'ясувалося, що це рішення дуже перенавантажує систему. Навіть за наявності шести-семи документів у бібліотеці це ускладнювало роботу системи. Це було зумовлено тим, що під час збереження оновленого документа для кожного знову викликався обробник події ItemUpdated. У свою чергу це викликало спробу оновлення всіх документів.

Цікаво те, що цей процес зупинявся через деякий час. Додавання перевірки оновлення (прапорця оновлення `wasUpdated`) показало, що навантаження на систему істотно не змінюється. Це дає підстави припустити, що під час виклику методу `SystemUpdate()` система Sharepoint перевіряє, чи змінювався документ. На жаль, код Sharepoint закритий, а офіційна документація не містить детальної інформації про роботу цього методу, тому неможливо це підтвердити. Прапорець оновлення залишено в коді для економії ресурсів за допомогою усунення зайвих викликів методу `SystemUpdate()` [3].

Під час роботи цього коду кількість запитів-порівнянь документів була $(n-1)^2$, де n — кількість документів у бібліотеці. Тобто складність такого алгоритму n^2 .

Запущені оновлення працювали одночасно, що призводило до перенавантаження системи.

Було розглянуто кілька варіантів вирішення цієї проблеми:

- додати в кожен тип вмісту прапорець «ще не оновлено». Під час оновлення документа синхронізувати та встановлювати прапорець. Якщо під час оновлення прапорець установлено, то тільки зняти його та не оновлювати документи. Цей варіант було відхилено, тому що після зняття прапорця знову викликатиметься оновлення документа, тобто такі зміни тільки збільшать навантаження системи;
- створити новий «основний» тип вмісту, який включає всі поля, і оновлювати тільки його під час оновлення документів. А під час оновлення «основного» типу вмісту оновлювати всі документи в бібліотеці;
- зробити те саме, що й у попередньому пункті, і додати окремий тип вмісту «прапорець» з полем «триває оновлення». Під час початку синхронізації «основного» типу з іншими встановлювати прапорець, після завершення знімати. Цей варіант було відхилено, тому що виклики методу `SystemUpdate()` синхронні й неможливо визначити, коли завершиться їх обробка.

Було вибрано варіант із додаванням «основного» типу вмісту як такий, що можна впровадити. Спочатку оновлюється основний документ, потім основний документ оновлює всі документи в бібліотеці крім себе, потім кожен документ намагається

оновити «основний». Тобто кількість запитів-порівнянь $1+(n-1)+(n-1)$. Це означає, що складність алгоритму дорівнює n .

Реалізація такого рішення має такий вигляд:

```
public override void ItemUpdated(SPItemEventProperties properties) // Елемент
додано
{
    SPListItem item = properties.ListItem; // Витягнення оновленого елемента
(документа)
    SPListItemCollection listItems = properties.ListItem.ListItems.List.Items; //
Вибір усіх елементів (документів) поточної бібліотеки (документів)
    Boolean updateDataStorageOnly = (item.ContentType.Name !=
"DataStorageContentType"); // Визначення, чи це основний документ
    foreach (SPListItem itemTmp in listItems) // Для всіх елементів (документів) у
поточній бібліотеці (документів)
    {
        if ((
            itemTmp.ID != item.ID) //Крім себе
            && (
                !updateDataStorageOnly
                || itemTmp.ContentType.Name == "DataStorageContentType" // Якщо
це не основний документ, то оновлюємо тільки основний документ
            )
        )
        {
            Boolean wasUpdated = false; // Прапорець оновлення
            foreach (SPField field in item.Fields) //Для кожного поля
            {
                if (field.Group == "Planning") // Поле, створене для ІС «Планування»
                {
```

```

        SPField fieldTmp =
itemTmp.Fields.GetFieldByInternalName(field.InternalName); // Спроба знайти таке саме
поле

        if (fieldTmp != null) // Поле знайдено
        {
            if (itemTmp[field.InternalName] != item[field.InternalName]) //
Поле має інше значення
            {
                itemTmp[field.InternalName] = item[field.InternalName]; //
Оновлення поля

                wasUpdated = true; // Оновлення прапорця оновлення
            }
        }
    }
}

if (wasUpdated) // Елемент (документ) оновлено
{
    itemTmp.SystemUpdate(false); // Збереження елемента (документа)
}
}
}
}

```

Висновки. Представлені у статті методика та засіб дозволяють автоматично синхронізувати документи в системі Sharepoint, що забезпечує більшу технологічність використання Sharepoint для проектування систем документообігу. Подібний підхід дозволив застосувати систему Sharepoint як платформу інформаційної системи «Планування наукових досліджень в Національній академії педагогічних наук України на базі мережі Інтернет».

Описані в статті труднощі розробки подібного засобу у Sharepoint й особливості роботи Sharepoint з документами можна використовувати як навчальний матеріал під час вивчення дисципліни «Системи електронного документообігу».

Список використаних джерел

1. *Jim Buyens*. Microsoft SharePoint Services Inside Out. — Microsoft Press, 2005. — 752 с.
2. Visual Studio 2008. — [Електрон. дані]. — Режим доступу: <http://msdn.microsoft.com/en-us/library/aa187919%28v=MSDN.10%29.aspx>. — Дата доступу: травень 2010. — Visual Studio 2008.
3. Windows SharePoint Services 3.0 SDK Documentation. — [Електрон. дані]. — Режим доступу: <http://msdn.microsoft.com/en-us/library/bb264594>. — Дата доступу: травень 2010. — SDK Documentation.

РАЗРАБОТКА СРЕДСТВ СИНХРОНИЗАЦИИ ДАННЫХ ДОКУМЕНТОВ НА ПЛАТФОРМЕ SHAREPOINT

Петрушко В. А.

Аннотация

Разработка системы автоматизации документооборота для управления образовательными процессами является актуальной задачей в условиях развития информационного общества. Одна из современных платформ для разработки таких систем — Microsoft SharePoint Products and Technologies. SharePoint содержит библиотеки для хранения документов, удобные средства для редактирования и управления версиями документов. В системах документооборота возникает потребность повторного использования данных в различных документах. SharePoint не предоставляет встроенные средства для синхронизации данных между документами. В статье описывается разработка средства для автоматической синхронизации данных между документами в SharePoint, на основе которого выполняется разработка информационной системы «Планирование научных исследований в Национальной академии педагогических наук Украины на базе сети Интернет».

Ключевые слова: синхронизация данных, документооборот, SharePoint, информационная система, Интернет, портал.

DEVELOPEMET OF DOCUMENT DATA SYNCRONIZATION TOOLS ON SHAREPOINT PLATFORM

Petrushko V.

Resume

The development of document management automation systems for managing by educational processes is an important task in the modern informational society. One of the popular platforms for such systems' development is Microsoft SharePoint Products and Technologies. SharePoint contains libraries for storing documents, convenient tools for editing and managing document versions. All workflow systems need to reuse data across different documents. SharePoint does not provide built-in tools for data synchronization across documents. This article describes the developement of the tool for automatic data synchronization across documents in SharePoint. This method is used as a basis for the Internet-based information system of the scientific researches planning in National Academy of Pedagogical Sciences of Ukraine.

Keywords: data syncronization, documents circulation, Sharepoint, information system, internet, portal.