

Для моніторингу публікацій я обрав [цей сайт](#). Насправді – публікації, це підписані документи(укази і т.д)

Які інструменти обрав:

1. ASP.NET WebAPI – клієнтська частина
2. EntityFramework 6 Code first approach
3. Log4net - бібліотека для логування
4. AngleSharp – бібліотека щоб парсити html
5. Castle Windsor – Inversion of Control

Результати:

Система складається з двох додатків, ASP.NET WebAPI і Console Application. Моніторинг сайту відбувається за рахунок TaskScheduler(a), який систематично запускає ConsoleApplication.

У WebAPI є метод, який дозволяє адміну триггернути логіку моніторингу вручну на заданий проміжок часу (наприклад запустити синхронізацію публікацій 12 Травня 2017 – 18 Травня 2017).

1. Business Logic:

Система сканує публікації сайту.

- Коли вона знаходить публікацію, якої немає в базі, то вона додає її в нашу систему.
- Якщо публікація була змінена на сайті, то в систему добавляється нова версія публікації, з тим самим Id, але з інкрементованим значенням Revision (стара публікація не видаляється). Це було реалізовано за рахунок Composite Primary Key, який складається з DocumentId і Revision
- Якщо публікація була видалена з сайту, то система відмічає в базі цю публікацію як видалену (IsDeleted = 1)

Завдяки цьому, в нашій системі завжди буде історія всіх публікацій, та видалені з сайту публікації.

2. Як я визначаю чи публікація була змінена?

Оскільки на [сайті](#) немає ніякої нормальної логіки щоб чітко ідентифікувати документ(публікацію), то я використовую комбінацію Title (Заголовок документа), HtmlContent (зміст самого документ), Link(Ссилка на документ).

В документа також є DocumentIdentifier (це номер підписаного документу), але я виявив, що не кожен документ має номер. Проте WebAPI має метод, щоб шукати документ за ідентифікатором

Тому в коді реалізовані інтерфейси `IEqualityComparer<Document>`, `IEquatable<Document>`, це дозволяє LINQ порівнювати Document Entity

Console Application

Може працювати в двох режимах (це визначається параметром який передається в exe)

`SearchNewDocumentsForLastWeek`, `SyncAllDocuments`

В першому режимі, програма синхронізує документи(публікації) сайту з базою за останій тиждень,
А в другому режимі, програма синхронізує абсолютно всі документи(публікації) сайту з базою.

WebAPI

Метод	Параметри	Опис
Document/GetDocumentByIdRecentRevision	documentId – id документу	Пошук документа по Id. Метод поверне останню версію документа
Document/GetDocumentRevision	documentId – id документу revision – номер версії документа	Пошук документа по Id і версії документа. Метод поверне один документ
Document/GetDocumentHistory	documentId – id документу	Пошук документа по Id. Метод поверне список версій одного і того же документа у порядку спадання версії
Document/GetDocuments	publishStartDate – ВІД час публішу публікації, publishEndDate – ДО час публішу публікації	Метод поверне список документі які були опубліковані в заданому проміжку
Document/GetDocumentByIdentifier	documentIdentifier – номер документа	Пошук документа за номером
Admin/SyncDocuments (в хедері або в queryString має бути admin Token)	publishStartDate – ВІД час публішу публікації, publishEndDate – ДО час публішу публікації	Синхронізує документи за вказаний проміжок часу