

Search or jump to...

[Pull requests](#) [Issues](#) [Marketplace](#) [Explore](#)

WSEI-cs201a / cs201-lab-Migracja-danych-via-XML-LinqToXML Public

[Watch](#) 1 [Fork](#) 7 [Star](#) 0

[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#)

master

1 branch

0 tags

[Go to file](#) [Add file](#) [Code](#)

kmolenda Update README.md

c7aa441 on 12 Dec 2019 4 commits

	.gitignore	Initial commit	3 years ago
	2019-0001.png	Initial upload	3 years ago
	2019-0001.xml	Initial upload	3 years ago
	LICENSE	Initial commit	3 years ago
	README.md	Update README.md	3 years ago
	issue.png	Initial upload	3 years ago
	issues.xml	Initial upload	3 years ago

README.md

# Migracja danych *via* XML

- Autor: Krzysztof Molenda
- Wersja: 1.0 (2019.12.11)

## Wprowadzenie

*Open Journal System* (w skrócie OJS) to system wspomagający prowadzenie czasopisma naukowego, realizując workflow: od zgłoszenia manuskryptu przez autorów, poprzez recenzje, opracowanie techniczne, aż po opublikowanie artykułu.

Oprogramowanie udostępniane jest za darmo i rozwijane na licencji GNU v.2.

Wielu wydawców korzysta z tego systemu. Do połowy 2018 r. obowiązywały wersje 2.x (zresztą nadal wspierane). Pod koniec 2018 r. pojawiła się - zupełnie zmieniona wersja OJS 3. Zmiany były fundamentalne, w wielu aspektach system został napisany ponownie, zmieniono nie tylko interfejs użytkownika, ale również architekturę bazy danych.

## Problem

Administratorzy opiekujący się systemem OJS 2.x napotkali poważne problemy przy migracji do OJS 3. Po wielu eksperymentach okazało się, że "najczystszy" sposobem jest zainstalowanie "świeżej" wersji OJS 3 i "ręczne" przenoszenie metadanych z wersji poprzedniej - na przykład za pomocą mechanizmów importu i eksportu *via* pliki XML.

Przedstawiony poniżej problem jest uproszczeniem i tylko fragmentem tego skomplikowanego procesu migracji.

## Zadanie

Dany masz plik `issue.xml` zawierający wyeksportowane z OJS 2.x metadane jednego z zeszytów czasopisma *Agricultural Engineering* - konkretnie vol. 21, issue 2.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE Issues SYSTEM "http://pkp.sfu.ca/ojs/dtds/2.4.7/native.dtd" PUBLIC "-//PKP/OJS Articles and Issues XML/EN">
<Issues>
  <Issue current="true" identification="num_vol_year" published="true">
    <volume>21</volume>
    <number>2(162)</number>
    <year>2017</year>
    <cover locale="en_US">
      <cover locale="pl_PL">
        <date_published>2017-07-26</date_published>
        <open_access/>
      </cover>
    </cover>
    <section>
      <title locale="en_US">Articles</title>
      <title locale="pl_PL">Artykuły</title>
      <abbrev locale="en_US">ART</abbrev>
      <abbrev locale="pl_PL">ART</abbrev>
      <article language="en" locale="en_US">
        <article language="en" locale="en_US">
          <article language="en" locale="en_US">
            <article language="en" locale="en_US">
              <article language="en" locale="en_US">
                <article language="en" locale="en_US">
                  <article language="en" locale="en_US">
                    <article language="en" locale="en_US">
                      <article language="en" locale="en_US">
                        <title locale="en_US">Geometrical Model of Lemon Fruit</title>
                        <title locale="pl_PL">Geometryczny model owoców cytryn</title>
                        <abstract locale="en_US">A proposal of a mathematical method of modelling of the lemon shape with Bézier's curves was presented. Lisbon, Verna, Genoa lemon cultivars were selected for verification of the modelling method. The lemon contour, which is its
```

```

meridian, was described with three smoothly combined Bézier's curves. Pictures taken in 10 locations changing every 360 were
the basis for description of lemon contours. Bézier's curves, which approximate meridians located on the surface of lemons, are
their 3-cem>D</em> models. The presented method may be applied for mathematical modelling of the lemon shape.</abstract>
<abstract locale="pl_PL">Przedstawiono propozycję metody matematycznego modelowania kształtu cytryn z wykorzystaniem
krzywych Béziera. Do weryfikacji metody modelowania wybrano cytryny odmian Lisbon, Verna, Genoa. Kontur cytryny, który
jest jej południkiem, opisano trzema gładko połączonymi krzywymi Béziera. Podstawą do opisu konturów cytryn są ich
fotografie wykonane w 10 zmieniających się co 360 położeniach. Krzywe Béziera aproksymujące południki leżące na
powierzchni cytryn są ich modelami 3<i>D</i>. Przedstawiona metoda może być stosowana do matematycznego modelowania
kształtu cytryn.</abstract>
</indexing>
<subject locale="en_US">Lemons; shape; Bezier's curves; mathematical models</subject>
<subject locale="pl_PL">cytryny; kształt; krzywe Béziera; modele matematyczne</subject>
</indexing>
<author primary_contact="true">
  <firstname>Leszek</firstname>
  <lastname>Mieszkalski</lastname>
  <affiliation locale="en_US">Institute of Production Organization and Engineering, Warsaw University of Life Sciences</affiliation>
  <country>PL</country>
  <email>mieszko@poczta.fm</email>
</author>
<pages>101-110</pages>
<date_published>2017-06-26</date_published>
</permissions>
<license_uri>http://creativecommons.org/licenses/by-nc-nd/4.0/</license_uri>
<copyright_holder locale="en_US">Agricultural Engineering</copyright_holder>
<copyright_holder locale="pl_PL">Agricultural Engineering</copyright_holder>
<copyright_year>2017</copyright_year>
</permissions>
<galley locale="en_US">
  <label>HTML</label>
  <file>
    <remote src="https://doi.org/10.1515/agriceng-2017-0020"/>
  </file>
</galley>
</article>
</section>
</issue>
</issues>

```

Dany masz przykład struktury pliku XML `2019-0001.xml`, który może zostać poprawnie zaimportowany do wersji OJS 3.

```

<?xml version="1.0"?>
<article xmlns:schemaLocation="http://pkp.sfu.ca/native.xsd" access_status="0" seq="1" section_ref="ART" date_published="2019-05-13" stage="production"
date_submitted="2019-12-03" locale="en_US" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://pkp.sfu.ca">
  <id advice="ignore" type="internal">20190001</id>
  <title locale="en_US">Influence of Drying Temperature and Moisture of Corn Seeds Planted on Their Damage</title>
  <prefix locale="en_US">The</prefix>
  <abstract locale="en_US">Evaluation of harvest and postharvest treatment quality of the grains is affected by usage of appropriate equipment and technology.
The grain damage depends on its physical and mechanical features. The corn planting has significant representation in European conditions, mainly for
food industry purposes. In the work, we were focused on observing of the drying temperature's impact on the micro-damage of the seeds. The solution of
the issue described in the work is closely related to the preparation and solution of the education project KEGA 039SPU-4/2017, and a scientific research
project VEGA 1/0718/17 „Study about the effect of technological parameters of the surface coating in agricultural and forestry techniques for qualitative
parameters, safety and environmental acceptability“</abstract>
  <licenseUri>https://creativecommons.org/licenses/by-nc-nd/4.0/</licenseUri>
  <keywords locale="en_US">
    <keyword>postharvest grain treatment</keyword>
    <keyword>micro-damage</keyword>
    <keyword>HACCP</keyword>
    <keyword>food industry corn</keyword>
  </keywords>
  <authors xsi:schemaLocation="http://pkp.sfu.ca/native.xsd" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
    <author user_group_ref="Author" include_in_browse="true">
      <givenname locale="en_US">Peter</givenname>
      <familyname locale="en_US">Bajus</familyname>
      <affiliation locale="en_US">Slovak University of Agriculture Nitra, Slovakia</affiliation>
      <country>SK</country>
      <email>noreply@wir.ptir.com</email>
    </author>
    <author user_group_ref="Author" include_in_browse="true">
      <givenname locale="en_US">Hiroslav</givenname>
      <familyname locale="en_US">Hraz</familyname>
      <affiliation locale="en_US">Slovak University of Agriculture Nitra, Slovakia</affiliation>
      <country>SK</country>
      <email>noreply@wir.ptir.com</email>
    </author>
    <author user_group_ref="Author" include_in_browse="true">
      <givenname locale="en_US">Ivan</givenname>
      <familyname locale="en_US">Rigo</familyname>
      <affiliation locale="en_US">Slovak University of Agriculture Nitra, Slovakia</affiliation>
      <country>SK</country>
      <email>noreply@wir.ptir.com</email>
    </author>
    <author user_group_ref="Author" include_in_browse="true" primary_contact="true">
      <givenname locale="en_US">Pavol</givenname>
      <familyname locale="en_US">Findura</familyname>
      <affiliation locale="en_US">Slovak University of Agriculture Nitra, Slovakia</affiliation>
      <country>SK</country>
      <email>pavol.findura@uniag.sk</email>
    </author>
    <author user_group_ref="Author" include_in_browse="true">
      <givenname locale="en_US">Adam</givenname>
      <familyname locale="en_US">Fürstentzeller</familyname>
      <affiliation locale="en_US">Slovak University of Agriculture Nitra, Slovakia</affiliation>
      <country>SK</country>
      <email>noreply@wir.ptir.com</email>
    </author>
    <author user_group_ref="Author" include_in_browse="true">
      <givenname locale="en_US">Pawel</givenname>
      <familyname locale="en_US">Kielbasa</familyname>
      <affiliation locale="en_US">Faculty of Production and Power Engineering Cracow, Poland</affiliation>
      <country>PL</country>
      <email>noreply@wir.ptir.com</email>
    </author>
    <author user_group_ref="Author" include_in_browse="true">
      <givenname locale="en_US">Urszula</givenname>
      <familyname locale="en_US">Malaga-Tobola</familyname>
      <affiliation locale="en_US">Faculty of Production and Power Engineering Cracow, Poland</affiliation>
      <country>PL</country>
      <email>noreply@wir.ptir.com</email>
    </author>
  </authors>
  <article_galley xsi:schemaLocation="http://pkp.sfu.ca/native.xsd" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" approved="false">
    <id advice="ignore" type="internal">1</id>
    <name locale="en_US">HTML</name>
    <seq>0</seq>
    <remote src="https://doi.org/10.1515/agriceng-2019-0001"/>
  </article_galley>
  <issue_identification>
    <volume>23</volume>
    <number>1</number>
    <year>2019</year>
  </issue_identification>
  <pages>5-12</pages>
</article>

```

Przykład pochodzi z innego zeszytu `vol. 23, issue 1` i uwzględnia nową strukturę metadanych przedstawioną w formacie XML. Przykład opisuje metadane pojedynczego artykułu.

Twoim zadaniem jest napisanie w C# aplikacji, która

1. wczyta plik `issue.xml` zawierający metainformacje związane z konkretnym zeszycem (n artykułów) w formacie eksportu z OJS 2.x,
2. we wskazanym folderze utworzy n plików zawierających metadane wszystkich artykułów, w formacie importu XML akceptowalnym przez OJS 3,
3. przy realizacji zadania wzorujesz się na przykładowych plikach XML.

Uwagi:

- W plikach wynikowych nie ma metainformacji związanych z językiem polskim (elementy z atrybutem `locale="pl_PL"`).
- Nazwy plików wynikowych to końcowy fragment numeru doi, odczytany z elementu `<remote src= ...>` dla danego artykułu. Na przykład dla `<remote src="https://doi.org/10.1515/agriceng-2019-0001"/>` nazwa pliku to `2019-0001.xml`

- Pewne informacje w nowym formacie są "sztywne" (np. odwołania do schematów `xmlns`). Należy po prostu je tam wprowadzić.
- W nowym formacie nowy element `<id type="internal" advice="ignore">20190001</id>` bezpośrednio w `<article>` (nie `<article_galley>`): jego wartość jest sklejeniem dwóch ostatnich kodów `doi:`.
- Część elementów i atrybutów w nowym formacie, to te same, które są w starym, ale ze zmienioną nazwą i w innej lokalizacji drzewa XML.
- W nowej wersji formatu XML tytuł artykułu rozbity jest na:
  - właściwy tytuł `<title>` oraz
  - `<prefix>` będący pierwszym wyrazem tytułu, który jest przedimkiem (w języku angielskim: *The, A lub An*)
 Przykład: tytuł jednego z artykułów `The Influence of Drying Temperature and Moisture of Corn Seeds Planted on Their Damage` powinien zostać rozbity na:

```
<title locale="en_US">Influence of Drying Temperature and Moisture of Corn Seeds Planted on Their Da
<prefix locale="en_US">The</prefix>
```

W systemach przetwarzania informacji bibliograficznych, ze względu na poprawność sortowania, stosuje się obecnie takie rozbijanie tytułów. (*The and other common words*)

## Cel ćwiczenia

Celem ćwiczenia jest - w oparciu o praktyczny problem - zapoznanie się z technikami przetwarzania dokumentów XML za pomocą bibliotek .NET, ze szczególnym uwzględnieniem *LINQ to XML*.

- LINQ to XML Overview (C#) @Microsoft
- Basic Queries (LINQ to XML) (C#) @Microsoft
- Learning LINQ Made Easy (LINQ to XML): Tutorial 3 @C# Corner

**UWAGA:** nie rozwiązujemy zadania za pomocą transformacji XLST.

## Referencje

Aby rozwiązać podany problem, nie ma potrzeby odwoływania się do poniższych referencji. Podane zostały z kronikarskiego obowiązku.

- Fragment oficjalnej dokumentacji systemu OJS 3 omawiający import metadanych o artykułach i zeszytach za pomocą XML: <https://docs.pkp.sfu.ca/importing-exporting/en/#articles-and-issues-xml-plugin>. Dokumentacja ta, niestety zawiera drobne błędy (niezmienione opisy z wersji 2.x).
- Projekt PKP-OJS na github.
- `native.xsd` - Schema describing native XML import/export elements specific to OJS.