

Phoenix Importer - dokument dla partnerów

Nasi partnerzy, chcąc zasilać Onet treściami, powinien mieć wystawione treści które chce oddawać w formacie Feed/ATOM ([https://pl.wikipedia.org/wiki/Atom_\(standard\)](https://pl.wikipedia.org/wiki/Atom_(standard))). Pełna specyfikacja standardu: <https://tools.ietf.org/html/rfc4287>.

ATOM jest wbudowany w wiele systemów zarządzania treścią (CMS), np:

- Wordpress: <https://wordpress.org/support/article/wordpress-feeds/> ,
- Joomla: https://docs.joomla.org/Help39:Extensions_Module_Manager_Syndication_Feeds.

Szczegóły implementacji

Poniżej wyszczególnione zostały pożądane cechy integracji:

1. Link do feeda musi być zabezpieczony przez certyfikat SSL (HTTPS) i mieć wyłączone cache (albo cache powinno być czyszczone przy każdym update);
2. Feed musi być zgodny ze standardem ATOM <https://tools.ietf.org/html/rfc4287>.
3. Dokument ATOM powinien mieć kodowanie UTF-8;
4. Dokument ATOM (zgodnie ze specyfikacją) musi zawierać pole <updated> dla całego feeda oraz <updated> przed każdym polem <entry>. Data w polu <updated> dla feeda musi być taka sama jak data <updated> dla ostatnio zaktualizowanego artykułu. Dzięki wartościom w tych polach system Onetu rozpozna, które treści ma pobierać/aktualizować;
5. Dokument ATOM musi być zgodny z formatem XML;
6. Artykuły muszą być posortowane od najnowszych do najstarszych po polu <updated> w <entry>. Wordpress sortuje po dacie pierwszej publikacji, wymagana jest zmiana w skórcie;
7. Zalecana ilość entry w feedzie, umożliwiającą m.in. poprawne sprawdzenie sortowania artykułów po dacie aktualizacji, to 10.
8. Jeśli artykułów wskazanych do eksportu jest więcej niż 50, feed powinien być stronicowany wg <https://tools.ietf.org/html/rfc5005#section-3>;
9. Każdy artykuł powinien w polu <entry> zawierać takie pola jak:
 - a. <id> - unikalny identyfikator artykułu, nie zmienny po aktualizacjach
 - b. <link> - link kanoniczny do artykułu
 - c. <title> - tytuł artykułu
 - d. <updated> - data aktualizacji w formacie zgodnym ze specyfikacją ATOM
 - e. <content> - treść artykułu w formacie html
 - f. <media:content> - wskazujący obrazek leadowy (główny). Podpis obrazka należy umieścić w polu <media:title>. Obrazek musi być podpisany. Przykład:

Przykład rozszerzenie media

```
<media:content url="https://domain.com/image.jpg">
  <media:thumbnail url="https://domain.com/image.jpg" type="image/jpeg"/>
  <media:credit role="author" scheme="urn:ebu">Image Author</media:credit>
  <media:title>Image title</media:title>
</media:content>
```

10. Dodatkowo wskazane jest, żeby w artykule <entry> były pola:
 - a. <author> - autor/autorzy artykułu
 - b. <source> - wskazuje źródło
 - c. <rights> - mówi o licencji
 - d. <category> - wskazuje przypisanie do kategorii. przykład:

Przykładowy tag

```
<category scheme="https://serwis.pl" term="testowy tag" />
```

11. Obrazki użyte w Atomie powinny mieć adresy absolutne (z <https://>). Adres do obrazka powinien wskazywać największy dostępny obrazek. Obsługiwane formaty obrazka to: png, jpg/jpeg, gif, webp. Obrazki będą importowane do systemów Onetu i skalowane w miarę potrzeb.
12. Pole <content> zawiera całą treść artykułu:
 - a. Paragrafy powinny być zawarte w tagach html <p>;
 - b. Importer treści wspiera następujące tagi html: html, body, div, p, h1, h2, h3, h4, h5, h6, b, strong, em, i, blockquote, ul, ol, li, iframe, embed, img, figure, figcaption, table, tbody, thead, th, tr, td;
 - c. Tagi niewspierane będą usunięte;
 - d. Usunięte będą atrybuty html (np style, class, id...);

- e. Wspierane video powinno być w formacie embed iframe lub w postaci linków do tych serwisów zawarte w znaczniku `<embed src="link-do-postu"></embed>`;
- f. Treści embedowane takie jak posty FB, Twitter'a oraz Instagram'a powinny mieć prawidłowe linki oraz być zamieszczone z użyciem znacznika `<embed> </embed>`;
- g. Podlinkowania (`<a>`) będą usunięte z treści artykułu;
- h. Blockquote będą interpretowane jako cytaty;
- i. Podpisy obrazków należy umieścić w atrybucie alt lub title. Każdy obrazek musi być podpisany:

Przykład podpis obrazka

```

```

- j. Pole `<content>` nie powinno przekraczać 255kB.

13. Pole `<summary>` jeśli występuje, nie powinno przekraczać 5kB. Długość pozostałych pól tekstowych nie powinna przekraczać 255 znaków.

Przykładowy dokument Atom

```
<?xml version="1.0" encoding="utf-8"?>
<feed xmlns="http://www.w3.org/2005/Atom"
  xmlns:media="http://search.yahoo.com/mrss/">
  <id>urn:uuid:2bc17842-15b0-4d97-9d5d-92ced783ee8</id>
  <link rel="alternate" type="text/html" href="https://serwis.pl"/>
  <link rel="self" type="application/atom+xml" href="https://serwis.pl/atom"/>
  <title>Testowy serwis</title>
  <updated>2019-07-29T10:22:39+02:00</updated>
  <entry>
    <id>urn:uuid:9bf2fffb-eda8-4735-8341-5bf33ab9d405</id>
    <link rel="alternate" type="text/html" href="https://serwis.pl/pilka-nozna/seo-lead-title/s4ny42r"/>
    <title>Tytu testowy</title>
    <updated>2019-07-29T10:22:39+02:00</updated>
    <published>2019-07-29T10:22:39+02:00</published>
    <rights>Testowa</rights>
    <source>
      <id>urn:uuid:98eblad5-d7fe-4f04-88ee-dd9f03a69f77</id>
      <title>testowe ródo</title>
      <updated>2019-07-29T10:22:39+02:00</updated>
    </source>
    <author>
      <name>jan kowalski</name>
    </author>
    <author>
      <name>joanna kowalska</name>
    </author>
    <category scheme="https://serwis.pl" term="testowy tag" />
    <category scheme="https://serwis.pl" term="pilne" />
    <summary>Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam molestie in sapien vel egestas.
Mauris sit amet
pellentesque ligula. Donec eget quam massa. Sed ut diam vestibulum, efficitur nisi sed, sollicitudin ipsum.
Donec id ligula nisl.
Duis aliquam tellus volutpat enim faucibus, ut euismod arc.</summary>
    <content type="html"><![CDATA[
      <p>Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam molestie in sapien vel egestas.
<strong>Mauris sit amet
pellentesque ligula.</strong> <em>Donec eget quam massa.</em> Sed ut diam vestibulum, efficitur nisi sed,
sollicitudin ipsum. Donec
id ligula nisl. Duis aliquam tellus volutpat enim faucibus, ut euismod arc.</p>
      <ol>
        <li>lista1</li>
        <li>lista2</li>
        <li>lista3</li>
      </ol>
      <p>Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam molestie in sapien vel egestas.
Mauris sit amet pellentesque
ligula.</p>
      
      <p>Donec eget quam massa. Sed ut diam vestibulum, efficitur nisi sed, sollicitudin ipsum. Donec id
```

```
ligula nisl. Duis aliquam
tellus volutpat enim faucibus, ut euismod arc.</p>
    <ul>
        <li>lista4</li>
        <li>lista5</li>
        <li>lista6</li>
    </ul>
    <h3>testowy ródtytu</h3>
    <p>Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam molestie in sapien vel egestas.
Mauris sit amet pellentesque
ligula.</p>
    <embed src="https://www.facebook.com/OnetWiadomosci/posts/10157822488758013"></embed>
    <iframe src="https://www.facebook.com/OnetWiadomosci/posts/10157822488758013"></iframe>
    <p>Donec eget quam massa. Sed ut diam vestibulum, efficitur nisi sed, sollicitudin ipsum. Donec id
ligula nisl. Duis aliquam
tellus volutpat enim faucibus, ut euismod arc. <a href="https://www.onet.pl/">https://www.onet.pl/</a></p>
    ]]></content>
    <media:content url="https://ocdn.eu/pulscms/MDA_/fa54337e466bb7e127795727a06efc3c.jpg">
        <media:credit role="author" scheme="urn:ebu">Jack Snow</media:credit>
        <media:title>sky light abstraction</media:title>
    </media:content>
</entry>
</feed>
```