

*ARTUR KULPA*

*JAKUB SWACHA*

*ROMAN BUDZOWSKI*

## **SUPERRANKINGI INTERNETOWE JAKO NARZĘDZIE WSPOMAGANIA DECYZJI KONSUMENTA**

### **1. Internetowe źródła informacji o produktach**

Konsument, dokonując zakupów przez Internet, musi się pogodzić z kilkoma ograniczeniami. Do najważniejszych z nich można zaliczyć brak fizycznego kontaktu z towarem oraz brak możliwości bezpośredniej konsultacji ze sprzedawcą. Ograniczenia te sprawiają, że osoba kupująca w sklepie internetowym jest zainteresowana znalezieniem dodatkowych informacji o poszukiwanym produkcie. W Internecie znajduje się wiele pomocnych źródeł informacji. Mogą być to materiały zamieszczone przez producenta czy sprzedawcę, recenzje produktów w profesjonalnych pismach lub serwisach internetowych, opinie o produktach na blogach, dyskusje na forach i listach mailingowych oraz różnego rodzaju rankingi.

Materiały, zamieszczone przez producenta czy sprzedawcę, zwykle ograniczają się do podania kategorii produktu i jego parametrów technicznych. Dla wielu grup produktów (na przykład książek, filmów samochodów, sprzętu elektronicznego i elektrotechnicznego) jest to zbyt mało, by zaspokoić potrzeby informacyjne konsumenta, który chce świadomie podjąć decyzje o zakupie. Czasami materiały takie zawierają sformułowania natury jakościowej, wtedy jednak mają charakter reklamowy i dalekie są od obiektywności.

Recenzje produktów w profesjonalnych pismach lub serwisach internetowych prezentowane są jako rzetelne źródła obiektywnej informacji. Niestety, nie zawsze tak jest w rzeczywistości. Często treść recenzji wynika z umów

reklamowych zawartych między producentem bądź dystrybutorem ocenianego produktu a wydawcą pisma lub serwisu. Ponadto, lista produktów, dla których dostępne są recenzje, zwykle tylko częściowo pokrywa się z listą produktów dostępnych na rynku.

Opinie o produktach wyrażone na blogach, choć subiektywne, bardzo często stanowią rzetelną informację o produkcie, jako że pisane są przez zadowolonych lub unieszczęśliwionych użytkowników. Niestety, dostępne są one dla bardzo niewielu produktów, ponadto trudno jest do nich dotrzeć.

Dyskusje na forach i listach mailingowych zwykle dotyczą dużej liczby produktów i zawierają wiele głosów. Niestety, bardzo często przybierają postać kłótni pomiędzy posiadaczami produktów różnych producentów, w której pojawia się niewiele argumentów merytorycznych, a jeśli nawet są, trudno jest postronnemu obserwatorowi ocenić ich wiarygodność.

Rankingi obejmują różne rodzaje produktów (na przykład samochody<sup>1</sup>, części komputerowe<sup>2</sup> czy gry komputerowe<sup>3</sup>) i mogą opierać się na wynikach sprzedaży<sup>4</sup>, opinii ekspertów<sup>5</sup> czy też opinii konsumentów<sup>6</sup>. Jako źródło informacji dla konsumenta, ranking ma formę oczywistą w interpretacji i łatwo pozwala porównać produkt z wyrobami konkurencyjnymi. Choć recenzja pojedynczego produktu zawiera bardziej szczegółowe informacje, brakuje jej wymowności rankingu, rzadko kiedy porównuje się w niej opisywany produkt z większą liczbą wyrobów konkurencyjnych.

Internet ułatwia przygotowanie i publikację rankingów – w konsekwencji można znaleźć ich w Internecie bardzo wiele. Rankingi sprzedaży czy popularności produktu prowadzone są przez większość sklepów internetowych, jak również przez niektóre serwisy tematyczne. Niestety, nie zawsze zachowują one obiektywność, bywa też, że są trudne do odnalezienia za pomocą wyszukiwarek. Do innych wad można zaliczyć ograniczony wybór produktów czy rzadki czas aktualizacji.

---

<sup>1</sup> <http://www.mojeauto.pl>.

<sup>2</sup> <http://rankingi.chip.pl>.

<sup>3</sup> <http://www.bigfishgames.com>.

<sup>4</sup> <http://sklep.proline.pl/shop.php?id=shopgaleria&lista=best>.

<sup>5</sup> <http://rankingi.chip.pl>.

<sup>6</sup> <http://www.mojeauto.pl>.

Wśród tak wielu rankingów, obarczonych wymienionymi wadami, rzeczą niezwykle pomocną dla konsumenta byłby superranking złożony z danych pochodzących z wielu tego typu publikacji dostępnych w Internecie.

Rankingi idealnie nadają się jako źródło danych do superrankingu, z uwagi na ich bezpośrednią kwantyfikowalność (pozycja produktu w rankingu jest liczbą). Zauważmy, że stworzenie zestawienia produktów opartego na treści na przykład list dyskusyjnych byłoby zadaniem o wiele bardziej skomplikowanym, gdyż wymagałoby zastosowania skomplikowanych technik przetwarzania języka naturalnego.

## **2. Koncepcja superrankingu jako narzędzia wspomaganie decyzji konsumenta**

Superranking to zestawienie sporządzone na podstawie zagregowanych danych pochodzących z wielu rankingów źródłowych. O pozycji produktu w superrankingu decydują jego pozycje w rankingach źródłowych skorygowane o wagi poszczególnych klasyfikacji.

Superranking ma za zadanie ukazać konsumentowi możliwie obiektywną informację o relatywnej atrakcyjności produktów w nim zawartych. Obiektywizację wyników uzyskuje się poprzez uwzględnienie w superrankingu jak największej liczby źródeł danych. Przy założeniu, że są to źródła niezależne, znacznie redukuje to znaczenie pojedynczych manipulacji, do jakich mogło dojść w rankingach źródłowych.

Superranking obejmuje szeroką listę produktów – wszystkie te, które znalazły się choć w jednym rankingu źródłowym. Dzięki temu konsument otrzymuje informację przetworzoną i gotową do wykorzystania przy podejmowaniu decyzji. Inaczej musiałby stracić wiele czasu na samodzielne wyszukiwanie i analizowanie zawartości rankingów źródłowych.

Niebagatelny wpływ na pozycję produktu w rankingu ma czas. Na rynku pojawia się wiele nowych towarów, a ich pozycja w rankingach może ulegać znacznym chwilowym wahaniom, będących wynikiem na przykład kampanii marketingowych bądź efektu „nowego produktu na rynku”. Superranking powinien nie tylko opierać się na aktualnych danych, ale również zapewniać możliwość przeglądania danych historycznych, pozwalając na analizę kształtowania się pozycji wybranego produktu w czasie, umożliwiając na przykład sprawdzenie najwyższej lokaty w superrankingu (bądź w innych rankingach) czy czasu

przebywania w czołówce produktów. Dzięki temu użytkownik mógłby się dowiedzieć, czy popularność produktu nie nosi znamion chwilowej mody.

Bardzo ważną zaletą superrankingu jest możliwość automatycznego generowania go (jak również automatycznej aktualizacji) na podstawie internetowych źródeł danych za pomocą języków skryptowych. W kolejnych punktach zostanie przedstawiony sposób implementacji superrankingu, jak również przykład jego praktycznego wykorzystania.

### **3. Sposób praktycznej implementacji generatora superrankingów**

Poniżej opisane zostaną podstawowe założenia systemu pozwalającego na automatyczne tworzenie superrankingów produktów na podstawie danych pochodzących z klasyfikacji umieszczanych na stronach internetowych.

Na początku zostaną scharakteryzowane wymagania funkcjonalne dotyczące takiego systemu oraz rozważone możliwe problemy, które mogą się pojawić przy jego konstrukcji wraz ze sposobami radzenia sobie z nimi. Następnie zostaną omówione techniczne zagadnienia związane z implementacją takiego systemu jako aplikacji internetowej.

Automatyczne tworzenie superrankingów oznacza, że system musi być zdolny, dla wskazanych źródeł:

- pobrać aktualne dane,
- przetworzyć kilka klasyfikacji w jeden wyjściowy superranking,
- zaprezentować otrzymany superranking.

Typowe źródło danych zawiera nazwę produktu (w zależności od rodzaju produktu mogą być ujęte dodatkowe oznaczenia, takie jak nazwa producenta lub autora) i aktualną pozycję; czasami mogą się znaleźć dodatkowe informacje, takie jak uzyskane punkty wpływające na pozycję w rankingu (na przykład wyniki sprzedaży, średnia ocen). Położenie tych danych powinno być jasno określone, nie tylko ze wskazaniem na odpowiedni dokument, ale również na odpowiedni znacznik, a nawet numer wiersza i słowa w przypadku niestrukturalizowanego tekstu czy słabo ustrukturalizowanych plików HTML. Dlatego w systemie możemy zidentyfikować dwa typy użytkowników. Pierwszy z nich: operator superrankingu, pełni funkcję administratora. Tylko taki użytkownik ma prawo dodać nowe źródła danych lub usunąć stare, określić częstotliwość aktualizacji czy czynniki wpływające na to, jakie znaczenie będzie miał każdy z rankingów. Dodawanie (lub modyfikowanie) źródła danych wymaga określo-

nej wiedzy informatycznej, pozwalającej na posługiwanie się wyrażeniami regularnymi do oznaczenia położenia danych rankingu w dokumencie źródłowym. W wielu przypadkach możliwe byłoby uproszczenie tego procesu (z punktu widzenia użytkownika, nie twórcy systemu) poprzez zastosowanie algorytmów ekstrakcji wiedzy bądź pozwolenie użytkownikowi na wskazanie myszką właściwego obszaru danych, niemniej jednak nie gwarantowałyby to poprawności działania tych rozwiązań, a znacząco wpłynęłyby na złożoność systemu, dlatego należy je traktować jako możliwe udoskonalenia.

Czytelnik klasyfikacji to klient, który chce poznać superranking dla wybranego rodzaju produktów. Ta czynność nie wymaga wiedzy informatycznej wykraczającej poza umiejętność posługiwania się prostym i intuicyjnym interfejsem użytkownika. Istotne mogłoby być, by czytelnik miał możliwość wpływania na niektóre z ustawień generatora (na przykład wykluczenia z rankingu wyników z wybranych źródeł czy też tymczasowej zmiany wag dla poszczególnych źródeł). Powinien mieć możliwość otrzymania superrankingu jednym kliknięciem myszy.

Czas ma kluczowe znaczenie w kontekście systemu. Rankingi źródłowe zmieniają się w czasie, tak samo musi zmieniać się wynikowy superranking. Powinien nie tylko automatycznie uwzględniać zmiany w rankingach, ale również czynniki wpływające na wagę poszczególnych klasyfikacji w superrankingu. Zazwyczaj źródłowe rankingi mają różną częstotliwość aktualizowania swoich zestawień. To od administratora systemu zależy, czy uznać za obowiązującą dla superrankingu najkrótszy z okresów (i wykorzystać procedurę okresowego obniżania wag starszych statystyk) czy też najdłuższy (i agregować dane z krótszych okresów w dłuższy). W przypadku rankingów *on-line*, które mogą zmieniać się co sekundę, administrator powinien decydować, jak często superranking ma się do nich dostosowywać (częste zmiany wymagają większej mocy obliczeniowej oraz przechowywania większej ilości danych).

Przetworzenie wielu rankingów źródłowych wymaga znormalizowania otrzymanych danych i ich zagregowania w jednym wynikowym superrankingu. Głównym powodem, dla którego normalizacja jest konieczna, jest fakt, że rankingi mogą się składać z różnej liczby pozycji. Jeżeli ranking źródłowy publikuje punkty określające kolejność w rankingu, normalizację można oprzeć właśnie na nich, w innym przypadku konieczne jest przyznanie każdej z pozycji określonej liczby punktów, im wyższa pozycja, tym większa liczba punktów. Całkowita punktacja każdego z rankingów źródłowych powinna być identyczna

– różnicowanie znaczenia poszczególnych rankingów odbywa się przy użyciu wag dopiero na etapie agregacji.

Normalizacja wiąże się również z tokenizacją nazw produktów, to jest ich zamianą tak, by każdy ujęty w rankingu produkt był jednoznacznie identyfikowany. Często ten sam produkt może występować pod różnymi nazwami (na przykład krótsze lub dłuższe nazwy modeli, nazwa produktu i jego wersja, różne nazwy w zależności od polityki marketingowej w danym regionie świata). W przypadku niektórych produktów jego różne wersje mogą występować jako pojedynczy wpis w rankingu lub jako kilka wpisów, w takim przypadku super-ranking powinien interpretować je jako pojedynczy wpis, sumując punkty z bardziej szczegółowych rankingów.

Agregowanie danych rozpoczyna się od wyznaczenia wagi dla poszczególnych rankingów, a następnie sumowane są ważone punkty rankingowe dla każdego z występujących w rankingach produktów. Wagi dla rankingów powinny uwzględniać ich wiarygodność (czy kryteria klasyfikacji są stosowne i obiektywne), obszerność (jak wiele produktów obejmował ranking) i popularność (ilu użytkowników uważa go za ważny). Wiarygodność może być określona zarówno na podstawie oceny operatora superrankingu, jak i czytelnika. Informacje o obszerności często można znaleźć na stronach rankingów źródłowych, zaczerpnięte stamtąd przez operatora superrankingu. Popularność klasyfikacji może być określona dzięki narzędziom internetowym monitorującym ruch w Internecie, takich jak Alexa.com<sup>7</sup>, Technorati<sup>8</sup> czy Gemius<sup>9</sup>, oraz na podstawie wskaźnika PageRank wyszukiwarki Google<sup>10</sup>. Dane te mogą być uzyskane w pełni automatycznie, dzięki czemu możliwe jest aktualizowanie wag rankingów bez wysiłku ze strony operatora superrankingu.

Należy zwrócić uwagę, że w przypadku trudności w wyznaczeniu właściwych wag można skorzystać z uproszczonej metody, przypisując wszystkim rankingom takie same wagi. W takim wypadku najlepiej jednak albo korzystać z dużej liczby rankingów źródłowych, albo uwzględniać tylko rankingi wysokiej jakości.

---

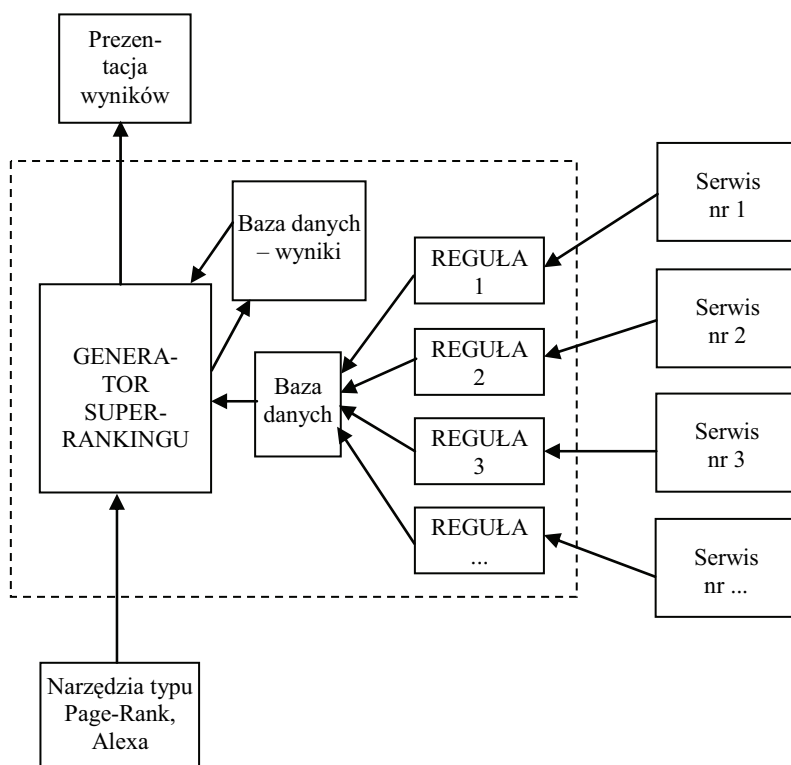
<sup>7</sup> <http://www.alexa.com>.

<sup>8</sup> <http://technorati.com/blogs/www.technorati.com?sub=bloginfo>.

<sup>9</sup> <http://www.ranking.pl>.

<sup>10</sup> L. Page, S. Brin, R. Motwani, T. Winograd, *The PageRank citation ranking: Bringing order to the Web*. Stanford Digital Library Technologies Project, <http://citeseer.ist.psu.edu/page98pagerank.html>, 1998.

Oprogramowanie automatycznie generujące superrankingi powinno wypełniać trzy podstawowe funkcje: pobranie, przetworzenie i prezentacja danych. Ogólny schemat takiego systemu przedstawiono na rysunku 1.



Rys. 1. Schemat systemu automatycznie generującego superrankingi

Źródło: opracowanie własne.

Dane wejściowe muszą być pozyskane z rankingów źródłowych i innych źródeł. Nie byłoby to konieczne, gdyby klasyfikacje były publikowane przy użyciu takich technologii, jak RSS, SOAP czy XML. W praktyce są one publikowane jako strony HTML.

System generujący superranking może być efektywnie zaimplementowany za pomocą języka Python<sup>11</sup>, który charakteryzuje bogate możliwości przetwa-

<sup>11</sup> C. Fehily, *Po prostu Python*, Helion, Gliwice 2002.

rzania danych, wbudowana obsługa protokołu HTTP i wyrażeń regularnych. W pierwszym etapie generowania superrankingu pobierane są rankingi źródłowe. Następnie dane dotyczące rankingów są wyodrębniane z dokumentu przy użyciu wyrażeń regularnych<sup>12</sup>, które można skonstruować dla dowolnego rankingów opartego na HTML. Jakkolwiek podejście strukturalne, oparte na modelu DOM<sup>13</sup>, byłoby bardziej eleganckie, nie może być użyte, gdyż dane rankingów zwykle nie występują w postaci wyodrębnionych elementów i muszą być wyodrębnione z innej zawartości.

Otrzymane nazwy produktów poddawane są tokenizacji. Wykorzystując słownik przygotowany przez operatora klasyfikacji, wszystkie nazwy synonimiczne zastępowane są jedną niepowtarzalną. Wszystkim produktom znajdującym się na liście zostają przypisane punkty rankingowe, zależnie od ich pozycji i rozmiaru rankingów. Następnie wszystkie nazwy wariantów produktu są usuwane i zastępowane nazwą ogólną, której przypisuje się sumę punktów rankingowych wszystkich wariantów.

W kolejnym etapie zostaje obliczona waga rankingów na podstawie stosowanej formuły, a punkty rankingowe wszystkich produktów na liście zostają przemnożone przez otrzymaną wagę. Rezultaty są przechowywane jako tymczasowy superranking; w trakcie przetwarzania kolejnych klasyfikacji źródłowych nowe elementy są dopisywane do listy, natomiast jeżeli nazwa produktu już się na niej znajduje, zwiększa się jedynie przypisaną jej liczbę punktów rankingowych.

Po przetworzeniu wszystkich rankingów źródłowych superranking jest porządkowany malejąco według punktów. Na końcu zostaje przetworzony do postaci dokumentu HTML i przekazany czytelnikowi rankingów.

Dane superrankingów przechowywane są w bazie danych, na przykład w systemie PostgreSQL<sup>14</sup>. W ten sposób istnieje możliwość analizy zebranych danych w dowolnym zakresie. Jako że poprzednie rankingi nie są kasowane, a jedynie oznaczane jako nieaktualne, możliwe jest przeprowadzenie analizy zmian rankingów w czasie.

---

<sup>12</sup> J. Friedl, *Mastering Regular Expressions*, O'Reilly, 1997.

<sup>13</sup> <http://www.w3.org/TR/2004/REC-DOM-Level-3-Core-20040407>.

<sup>14</sup> J.C. Worsley, J.D. Drake, *Practical PostgreSQL*, O'Reilly 2002.



#### 4. Przykład wykorzystania superrankingu

Aby ukazać przykład funkcjonowania superrankingu, stworzono uproszczoną wersję generatora superrankingów, wykorzystującego rankingi filmów fabularnych z witryn internetowych: [www.kinomaniak.pl](http://www.kinomaniak.pl), [www.ranking.filmow.dvd.ys.pl](http://www.ranking.filmow.dvd.ys.pl) i [www.filmweb.pl](http://www.filmweb.pl).

Każda z tych witryn prowadzi własny ranking filmów obejmujący zróżnicowaną liczbę pozycji: [www.kinomaniak.pl](http://www.kinomaniak.pl) i [www.filmweb.pl](http://www.filmweb.pl) mają po 100 pozycji, natomiast [www.ranking.filmow.dvd.ys.pl](http://www.ranking.filmow.dvd.ys.pl) – 20 pozycji. Każda z tych klasyfikacji podaje liczbę głosów przypisanych kolejnemu tytułowi filmowemu.

Do wygenerowania superrankingu posłużono się programem napisanym w języku Python<sup>15</sup>. Najpierw pobrane zostały z Internetu dokumenty zawierające rankingi źródłowe, a następnie wyłuskano z nich dane potrzebne do budowy superrankingu, to jest pozycję filmu w rankingu, tytuł oraz liczbę przypisanych mu głosów.

Dane te zostały pobrane z witryny internetowej za pomocą mechanizmu wyrażeń regularnych zastosowanego na dokumentach HTML zawierających rankingi źródłowe. Dla każdego rankingi źródłowego zostały zbudowane osobne wyrażenia regularne:

##### 1. [www.kinomaniak.pl](http://www.kinomaniak.pl)

```
<tr><td bgcolor="#[\w]{6}" width="20"><font class ="male"
><font color="black"></font></font><center><font class="ma-
le"><font color="black">([0-9]{1,3})</font></font></ cen-
ter></td><td bgcolor="#[\w]{6}" width="330"><font class=
"male"><a class="piaty"
href="http://www\.\kinomaniak\.\pl/filmy/[^"]+"><b>([^\<]+) (<b
r>)?</b>[^<]*</a></font></td><td bgcolor="#[\w]{6}"width=
"20"><font class="male"><font color="black"></font></font
><center><font class="male"><font color="black">[0-9]{4}</
font></font></center></td><td bgcolor="#[\w]{6}" width="80"
><font class="male"><font color="black"></font></font><
center><font class="male"><font color="black">([0-9]+\.[0-9]
{2})</font></font></center></td><td bgcolor="#[\w]{6}" wid-
```

<sup>15</sup> C. Fehily, *Po prostu Python...*

```
th="80"><font class="male"><font color="black"></font></font><center><font class="male"><font color="black">([0-9]+)</font></font></center></td></tr>
```

### 2. www.ranking.filmow.dvd.ys.pl

```
<tr onclick="location\.'" bgcolor="^[^"]+">\n<td align="center"><font face="Verdana" size="2">([0-9]{1,2})</font></td>\n<td align="center"><font face="Verdana" size="2">([0-9]+)</font></td>\n<td align="center"><font face="Verdana" size="2">([0-9]+\.[0-9]{3})</font></td><td><a class="film" href="http://www\.ranking\.filmow\.dvd\.ys\.pl/index\.php\?strona=opis_filmu&amp;choice=7&amp;id_filmu=[0-9]{1,4}">([^\<]+)</a></td></tr>
```

### 3. www.filmweb.pl

```
<tr class="gr[0-9]">[\s]*<td class="p12">([0-9]{1,3})</td>[\s]*<td>[\s]* [0-9]{0,2}</td>[\s]*<td class="p13"><a name="^[^"]+">[\s]*</a><a class="colorUnderline" title="^[^"]+" - o filmie - FILMWEB\.pl" href="http://^[^"]+">[\s]*)?[\s]*<b>([^\<]+)</b><br>)?[\s]*[^\<]+</a>[\s]*</td>[\s]*<td class="p11">([0-9],[0-9]{3})</td>[\s]*<td class="p11">([0-9]{1,6})</td>[\s]*</tr>
```

Tak uzyskane dane zostały poddane tokenizacji. Wykorzystano przy tym prosty, heurystyczny mechanizm sprowadzania alternatywnych wersji tytułów do wersji najkrótszej, poprzez usuwanie „wtreć” o ograniczonej długości (na przykład *Gwiezdne wojny część 1: Nowa nadzieja* i *Gwiezdne wojny: Nowa nadzieja* sprowadzono do tej drugiej wymienionej postaci). Tak zbudowana została lista niepowtarzalnych 205 tytułów filmów.

Dla każdego z filmów ujętych w superrankingu obliczono następnie syntetyczny wskaźnik ważonej pozycji w rankingu ( $r$ ), korzystając ze wzoru:

$$r(n) = P_{\max} - \frac{\sum_{i=1}^3 p_i(n)(1-w_i)}{3},$$

gdzie:

- $n$  – jednoznaczny numer filmu,
- $p_i(n)$  – pozycja  $n$ -tego filmu w  $i$ -tym rankingu,
- $P_{\max}$  – numer maksymalnej pozycji we wszystkich rankingach,
- $w_i$  – suma głosów w  $i$ -tym rankingu/suma głosów wszystkich rankingów.

Następnie lista filmów, ujęta w superrankingu, została uporządkowana od filmu o największym wskaźniku ważonej pozycji w rankingu do filmu o najmniejszym wskaźniku. W końcowym etapie wygenerowany został dokument HTML prezentujący superranking w postaci pokazanej na rysunku 2.

## Zestawienie superrankingu filmów fabularnych

Wybrane serwisy z rankingami filmów:

1. [www.kinomaniak.pl](http://www.kinomaniak.pl)
2. [www.ranking.filmow.dvd.ys.pl](http://www.ranking.filmow.dvd.ys.pl)
3. [www.filmweb.pl](http://www.filmweb.pl)

Pozycja	Nazwa filmu	Uzyskany wskaźnik
1.	Ojciec chrzestny	99.9419629468
2.	Ojciec chrzestny II	99.8839258936
3.	Skazani na Shawshank	99.8258888404
4.	Dawno temu w Ameryce	99.7678517873
5.	Człowiek z blizną	99.7098147341
6.	Lot nad kukułczym gniazdem	99.6517776809
7.	Milczenie owiec	99.5937406277
8.	Pewnego razu na Dzikim Zachodzie	99.5357035745
9.	Forrest Gump	99.4776665213
10.	Mechaniczna pomarańcza	99.4196294681
11.	Obywatel Kane	99.361592415
12.	Pulp Fiction	99.3035553618

Rys. 2. Widok raportu superrankingu

## Podsumowanie

W pracy przedstawiono schemat systemu pozwalającego na automatyczne tworzenie superrankingów produktów na podstawie danych pochodzących z rankingów umieszczanych na stronach internetowych i zaproponowano jego wykorzystanie jako narzędzia wspomagania decyzji konsumenta.

Opisano podstawowe zasady działania systemu, jego wymogi funkcjonalne i sposoby rozwiązywania typowych problemów, a także prostą przykładową implementację systemu przetwarzającą rankingi popularności filmów.

W przekonaniu autorów, superranking stanowi bardzo poręczne narzędzie dla konsumentów. Dzięki wykorzystaniu wielu źródeł jest w znacznie mniejszym stopniu podatne na nierzetelność wynikającą ze stronniczości wyrażanych opinii, niż każde z tych źródeł z osobna, a także pozwala w łatwy sposób porównać między sobą szeroki zakres produktów. Należy się zatem spodziewać, że superrankingi znajdą liczne zastosowania praktyczne.

## Literatura

Alexa, <http://www.alexa.com>, 2007.

Big Fish Games, <http://www.bigfishgames.com/>, 2007.

Chip.pl, <http://rankingi.chip.pl>, 2007.

Fehily C., *Po prostu Python*, Helion, Gliwice 2002.

Friedl J., *Mastering Regular Expressions*, O'Reilly, 1997.

Page L., Brin S., Motwani R., Winograd T., *The PageRank citation ranking: Bringing order to the Web. Stanford Digital Library Technologies Project*, <http://citeseer.ist.psu.edu/page98pagerank.html>, 1998.

Mojeauto.pl, <http://www.mojeauto.pl>, 2007.

ProLine, <http://sklep.proline.pl/shop.php?id=shopgaleria&lista=best>, 2007.

Ranking.pl, <http://www.ranking.pl>, 2007.

Technorati, <http://technorati.com/blogs/www.technorati.com?sub=bloginfo>, 2007.

W3C, *Document Object Model (DOM) Level 3 Core Specification*, <http://www.w3.org/TR/2004/REC-DOM-Level-3-Core-20040407>, 2004.

Worsley J.C., Drake J.D., *Practical PostgreSQL*, O'Reilly, 2002.

**INTERNET SUPER-RANKINGS  
AS CONSUMER DECISION SUPPORT TOOL**

**Summary**

When buying in the Internet, consumer lacks both the opportunity to physically test the product and to consult the seller immediately. If he or she needs more information to make the decision, there is plenty of information sources in the Internet, where he or she can find out how the product is evaluated by experts and other customers. However, there are many flaws of such sources, such as being hard to find, low credibility, and often, a small number of compared products.

In this paper we propose that a super-ranking based on an input set of source product rankings can supply more usable information and thus can be used as consumer decision support tool. We provide a detailed description of how a super-ranking can be automatically generated and present a sample super-ranking of popular films.

*Translated by Jakub Swacha*

