

# Ilona Pawełoszek-Korek

---

## Praktyczne aspekty pozyskiwania danych z portalu aukcyjnego w celu ekstrakcji wiedzy

---

Ekonomiczne Problemy Usług nr 57, 603-611

---

2010

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej [bazhum.muzhp.pl](http://bazhum.muzhp.pl), gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

ILONA PAWEŁOSZEK-KOREK

Politechnika Częstochowska

ipaweloszek@zim.pcz.pl

## PRAKTYCZNE ASPEKTY POZYSKIWANIA DANYCH Z PORTALU AUKCYJNEGO W CELU EKSTRAKЦИИ WIEDZY

### Wprowadzenie

Aukcje internetowe jako model e-biznesu zyskują sobie coraz większą popularność wśród przedsiębiorstw różnych i branż i wielkości. W ostatnich latach wiele badań i publikacji poświęcono różnym aspektom aukcji internetowych. Wiele wysiłków badawczych skupia się na wykrywaniu nadużyć i oszustw<sup>1 2</sup>. Ważnym polem badawczym jest także zastosowanie podejścia agentowego do modelowania, projektowania i implementacji internetowych platform aukcyjnych<sup>3</sup>. Opracowywane są także systematyczne podejścia do badania aukcji pod kątem doświadczeń

---

<sup>1</sup> J. Mena, *Investigative Data Mining for Security and Criminal Detection*, Butterworth-Heinemann, Boston 2003.

<sup>2</sup> Y. Ku, Y. Chen, C. Chiu, *A Proposed Data Mining Approach for Internet Auction Fraud Detection*, in: *Intelligence and Security Informatics*. pp. 238-243. Springer Berlin, Heidelberg 2007.

<sup>3</sup> I. Jureta, M. Kolp, S. Faulkner, *An Agent-Oriented Perspective on E-bidding Systems*, in: Vijay Sugumara (red.), *Intelligent Information Technologies and Applications*, pp 94-118. IGI Publishing, 2007.

klienta<sup>4 5</sup>. Portale aukcyjne konkurują ze sobą nawzajem, przedmiotem badań są zatem także porównania oparte na analizie praktycznych rozwiązań<sup>6</sup>.

Celem niniejszego artykułu jest zaprezentowanie zaprojektowanego przez autorkę praktycznego rozwiązania, służącego do ekstrakowania danych z portali aukcyjnych w celu pozyskania z nich wiedzy, która może być przydatna przedsiębiorstwu prowadzącemu biznes za pomocą aukcji internetowych. Zaproponowane narzędzie o nazwie AuMiner jest elastycznym i funkcjonalnym rozwiązaniem opartym na usługach sieci Web. Ma za zadanie pozyskiwać dane o określonej strukturze w ustalonych odstępach czasu i eksportować je w formacie wygodnym do dalszego przetwarzania.

## 1. Platformy aukcyjne jako środowisko biznesu

Główną zaletą portali aukcyjnych jest to, iż pozwalają one kupującym znaleźć w jednym miejscu szeroką gamę towarów o zróżnicowanych cenach, a sprzedającym umożliwiają dotarcie do tysięcy potencjalnych klientów po relatywnie niskim koszcie. Platformy aukcyjne z roku na rok zyskują popularność wśród przedsiębiorców, którzy wybierają je jako swoje podstawowe środowisko biznesowe i źródło dochodu. Ze względu na niski koszt wejścia na rynek model e-aukcji stwarza idealne możliwości biznesowe dla małych firm handlowych. Aukcje typu B2C stanowią nowe kanały dystrybucji. Natomiast w relacjach B2B ułatwiają one kooperację przez redukcję kosztów transakcji i wyszukiwania partnerów biznesowych, rozmywają granice pomiędzy firmami i zmieniają strukturę aliansów, które istniały w przeszłości<sup>7</sup>.

Współczesne portale aukcyjne są skomplikowanymi systemami, które dostarczają użytkownikom funkcjonalnych i przyjaznych interfejsów. Spośród wielu elementów, których jakość tworzy ogólny wizerunek portalu, na uwagę użytkowników biznesowych szczególnie zasługują interfejsy Web API. Nowoczesne platformy aukcyjne poprzez ten interfejs udostępniają usługi dostępu do swoich baz danych, co znacznie podnosi atrakcyjność portalu jako partnera biznesowego, gdyż umożliwia jego integrację z zewnętrznymi systemami informatycznymi.

---

<sup>4</sup> A. J. Haywood, *Online Auctions: User Experience Insights from eBay*, Chimera Working Paper, Colchester: University of Essex 2006.

<sup>5</sup> D.H. Parente (red.), *Best Practices for Online Procurement Auctions*, IGI Global, USA 2008.

<sup>6</sup> W. Chmielarz, *Próba analizy porównawczej serwisów aukcji internetowych w Polsce*, w: *Roczniki Kolegium Analiz Ekonomicznych vol.16/2006*, s.13-26, Wydawnictwo SGH, Warszawa 2006.

<sup>7</sup> Pinker E., Seidmann A., Vakrat Y., *Managing online auctions: Current business and research issues*. Management Science, 49(11), s. 1457-1484, 2003 s. 1457.

## **2. Problemy pozyskiwania wiedzy o otoczeniu**

Badanie otoczenia przedsiębiorstwa jest ważne dla każdej organizacji, zarówno szukającej trwałej przewagi konkurencyjnej, jak i realizującej przedsięwzięcie mające przynieść jednorazowy zysk. W przypadku biznesu prowadzonego online badania użytkowników mają zwykle trzy główne cele: ustalenie potrzeb klienta, odkrycie trendów, zależności, wzorców w zachowaniu klientów oraz monitorowanie działań konkurencji.

W przypadku aukcji online bardzo trudno jest zastosować bezpośrednio metody badawcze, takie jak ankiety czy wywiady skierowane do użytkowników portalu. Wiele portali ukrywa dane kontaktowe osób biorących udział w aukcjach, takie działanie ma na celu zabezpieczenie przed oszustwami. Ponadto masowe wysyłanie wiadomości do użytkowników mogłoby być potraktowane jako spam. Zadawanie pytań na forach dyskusyjnych może być mało efektywne, ponieważ korzysta z nich tylko niewielki procent osób.

Zatem najodpowiedniejsze są pośrednie metody badań, takie jak regularna obserwacja, zbieranie danych z aukcji i analizowanie ich metodami statystycznymi, aby wyodrębnić użyteczną wiedzę. Każdego dnia w portalach aukcyjnych dokonywane są tysiące transakcji. Zbieranie danych do analizy w sposób ręczny jest bardzo uciążliwe, wymaga bowiem przeglądania zakończonych aukcji i notowania najważniejszych faktów. Takie postępowanie możliwe jest tylko dla wąskich kategorii produktów lub kilku wybranych sprzedawców. Eksploracja danych z zakończonych aukcji internetowych może pozwolić odkryć wiele nieznanych dotąd faktów i wskazać szanse na nowe, rentowne przedsięwzięcia. Efekty badania i użyteczność wyników są w dużym stopniu uzależnione od jakości pozyskanych danych, ich kompletności i właściwego formatu.

## **3. Prezentacja narzędzia AuMiner**

AuMiner jest narzędziem do pozyskiwania danych z portali aukcyjnych w celu ekstrakcji wiedzy w postaci trendów, wzorców reguł i zależności opisujących rynek aukcyjny lub jego wycinek. Wiedza ta pozwoli sprzedającym kształtować strategię sprzedaży, tak aby poprawić swoją konkurencyjność. Główną funkcją aplikacji AuMiner jest okresowe zbieranie danych z zakończonych aukcji w sposób zautomatyzowany i eksportowanie ich do systemów eksploracji danych. Narzędzie AuMiner jest łatwo konfigurowalne, a dane zebrane za jego pomocą stanowią wysokiej jakości surowiec do analiz statystycznych. Aplikacja jest obecnie w fazie testowej, współpracuje z interfejsem Web API portalu aukcyjnego Allegro<sup>8</sup>. Śro-

---

<sup>8</sup> <http://webapi.allegro.pl/uploader.php> (listopad 2009).

dowisko testowe to: serwer Apache z bazą danych MySQL 5.0.51 i PHP 5.2.5. Ta konfiguracja posiada liczne zalety, do których zaliczyć można: dostępność różnorodnych bibliotek programistycznych, możliwość połączenia z nowoczesnymi językami do tworzenia interfejsów, takimi jak XHTML i Ajax. Dodatkowym argumentem przemawiającym za zastosowaniem języka PHP5 było jego rozszerzenie SOAP, które w prosty sposób pozwala tworzyć aplikacje korzystające z usług sieci Web, jak również publikować własne usługi.

Baza danych MySQL została wykorzystana do przechowywania danych aukcyjnych pobranych przez AuMiner oraz specyfikacji WSDL portalu Allegro. Baza posiada możliwość eksportu danych do różnych formatów, takich jak: CSV, MS Excel, LaTeX, Open Document, PDF, SQL, XML. Aby skorzystać z usług Web API, należy posiadać konto w portalu Allegro i klucz użytkownika, który można uzyskać po wysłaniu prośby do administratorów portalu pocztą elektroniczną. Zgodnie z praktyką powszechnie stosowaną w rozwiązaniach hurtowni danych AuMiner zapisuje dane z systemu źródłowego w tabelach bazy danych o płaskiej strukturze. Struktura taka zapewnia najlepszą wydajność. Importowanie lub eksportowanie z pliku o płaskiej strukturze jest prawdopodobnie najszybszą metodą w porównaniu z zastosowaniem innych formatów (np. XML)<sup>9</sup>.

Pierwszy krok pracy z narzędziem AuMiner to sformułowanie zapytania. Należy ustalić pożądaną strukturę danych oraz warunków ograniczających wyniki wyszukiwania (np. do określonej kategorii przedmiotów aukcji, typów aukcji, użytkownika wystawiającego aukcje, daty). Decyzja o strukturze danych i ograniczeniach wyszukiwania powinna być podyktowana celem badania – należy postawić sobie hipotezy i zastanowić się, jakie dane są niezbędne do ich udowodnienia bądź odrzucenia. Sprzedawcy korzystający z portalu aukcyjnego mogą się skoncentrować, przykładowo, na pozyskaniu następujących informacji, które pomogą im skutecznie konkurować:

1. Najlepsze praktyki sprzedawców aukcyjnych: W jakich dniach i godzinach uzyskiwane są najwyższe ceny w licytacjach? Czy publikowanie aukcji jako promowanych bądź wyróżnionych rzeczywiście daje efekty w postaci częstszych zakupów? Czy publikowanie aukcji z miniaturką (zdjęciem w wynikach wyszukiwania) przyciąga większą liczbę klientów?
2. Popularność produktów i zachowanie klientów: Jaki typ aukcji daje największe dochody: licytacja czy stała cena? W jakiej kategorii produktów uzyskiwane są największe przychody? Jaki sezon jest najlepszy do wystawienia aukcji określonego produktu? Jaki produkt jest najczęściej kupowany w danej kategorii?

---

<sup>9</sup> V. Rainardi, *Building a data warehouse, with examples in SQL server*. Apress USA 2008, s. 187.

3. Konkurencja: Kto jest najsilniejszym konkurentem, sprzedającym najwięcej produktów lub uzyskującym największą wartość sprzedaży? Z jakich miast i regionów są konkurenci i ilu sprzedawców z określonego regionu oferuje ten sam produkt? Czy istnieje zależność pomiędzy geograficzną lokalizacją sprzedawcy a ceną uzyskiwaną w aukcji lub ilością sprzedanych produktów?
4. Rynek: Jakie kategorie produktów charakteryzują się sezonowością? Jak kształtuje się dynamika sprzedaży? Czy istnieje korelacja pomiędzy dwoma produktami lub kategoriami?

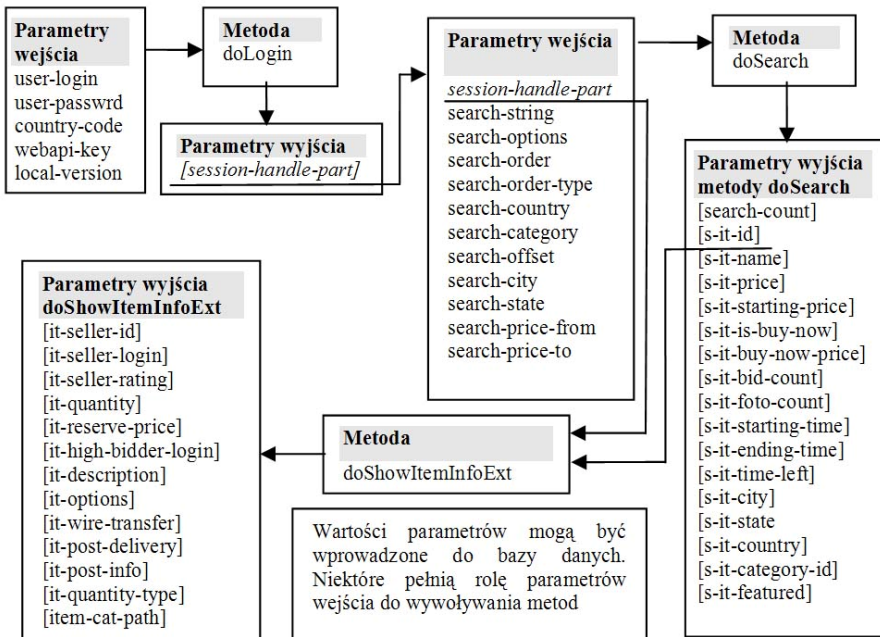
Kolejnym krokiem jest wprowadzenie zapytania za pomocą formularza. Interfejsem użytkownika AuMiner jest przeglądarka internetowa. Po sformułowaniu zapytania i zatwierdzeniu go dalsze zadania są realizowane przez agenta programowego, który odbiera parametry zapytania w postaci tablicy, a następnie przeszukuje bazę dostępnych metod Web API (a konkretnie ich parametrów wyjściowych) pod kątem znalezienia tych, które są w stanie zwrócić dane żądane przez użytkownika. Dla poprawienia szybkości przeszukiwania specyfikacja WSDL została wprowadzona do relacyjnej bazy danych dostępnej lokalnie. Agent wywołuje metody w sposób rekurencyjny: najpierw stara się wywołać metodę zwracającą wartości końcowe, jeśli nie może tego zrobić z powodu braku niektórych danych wejściowych, to szuka metody, która może je dostarczyć. Przykład kolejności wywołań metod zaprezentowano na rysunku 1.

Na szczególną uwagę zasługują metody:

- doSearch – pobiera jako parametry kryteria wyszukiwania, a zwraca spełniające je rekordy (w porcjach po 50 – zatem jeśli ogólna liczba wyników wyszukiwania jest większa niż 50, to procedura jest uruchamiana przez agenta więcej niż 1 raz),
- doShowItemInfoExt – metoda musi być wywołana w pętli dla każdego rekordu zwróconego przez metodę doSearch, zwraca szczegóły każdej znalezionej aukcji i login użytkownika – sprzedawcy i kupującego, które to dane mogą być użyte do późniejszego wywołania metody doShowUser podającej dokładne informacje o użytkownikach.

Przed wprowadzeniem danych do bazy MySQL niektóre z nich powinny być przekonwertowane do formatu bardziej wygodnego do analizy i wizualizacji (np. czasy zakończenia i rozpoczęcia aukcji są podane jako wartości liczbowe typu integer). Opcje przedmiotu aukcji (takie jak miniaturka, rodzaj transportu i płatności) są tablicą wartości typu integer, zatem powinny być konwertowane do bardziej czytelnego formatu (np. opcje transportu i płatności mogą być przetłumaczone na zestaw pewnych predefiniowanych wartości: list polecony, paczka, odbiór osobisty, płatność z góry, za pobraniem itd.). Aby uzyskać odpowiednią ilość danych pozwalającą na zastosowanie technik eksploracji, należy zaprogramować AuMiner, aby pobierał i zapisywał dane okresowo. Częstotliwość pobierania danych może być za-

leżna od tego, jak długo portal aukcyjny przechowuje historię aukcji, w przypadku Allegro jest to 1 miesiąc. Częstotliwość może zależeć także od specyfiki przedmiotów aukcji.



Rys. 1. Przykładowa kolejność wywołań usług

Źródło: I. Pawełoszek-Korek, *A Framework for Online Auctions Market Research*, w: *Proceedings of BIR'2008 the Seventh International Conference on Perspectives in Business Informatics Research*, Ed. Stanisław Wrycza, Gdańsk 2008, s. 306.

#### 4. Pozyskiwanie wiedzy z danych

Dane są jedynie surowcem, który potrzebuje dalszej obróbki, aby stać się produktem finalnym – informacją lub wiedzą. Innymi słowy, informacja jest to znaczenie, jakie zyskują dane dzięki sposobowi ich interpretacji<sup>10</sup>. Metody interpretacji zależą od liczby rekordów w zbiorze danych. Mając do czynienia z dużymi zbiorami danych, na pierwszy rzut oka nie da się ustalić żadnych faktów czy reguł. W przypadku analizy danych aukcyjnych można mieć do czynienia z tysiącami lub

<sup>10</sup> Collins *New compact English dictionary*, 2nd ed, Harper-Collins, Edinburgh 1992.

dziesiątkami tysięcy rekordów zależnie od horyzontu czasowego analizy i badanej kategorii aukcji.

W przypadku bardziej ogólnych obserwacji (na przykład kategoria TV, wideo i DVD) liczba rekordów może dochodzić do kilku tysięcy miesięcznie. Taka liczba danych determinuje potrzebę zastosowania oprogramowania statystycznego lub narzędzi Data Mining. Zadania, jakie powinno realizować wybrane oprogramowanie analityczne, to przede wszystkim:

- opisywanie i podsumowanie danych tabelarycznych: obliczanie częstości występowania, dystrybucji, procentów, podsumowanie w oparciu o wybrane atrybuty,
- testowanie hipotez,
- eksploracja danych w celu wyodrębnienia struktur, wzorców lub klastrów: obserwacja trendów, analiza klastrów, drzewa klasyfikacyjne,
- poszukiwanie relacji: korelacja i regresja.

Wiele narzędzi statystycznych przeznaczonych do eksploracji danych jest dostępnych zarówno komercyjnie, jak i na licencji Open Source. Do aplikacji Open Source można zaliczyć: program Weka, język R i Orange. Przykłady komercyjnych narzędzi to SPSS Clementine czy Statistica.

## **Podsumowanie**

Aukcje online stały się popularnym sposobem prowadzenia biznesu, platformy aukcyjne stanowią środowisko, w którym współdziałają firmy i osoby prywatne zarówno jako sprzedawcy, jak i kupujący. Dlatego badania i narzędzia ułatwiające uczestnikom rynku aukcyjnego zdobywanie wiedzy o kierujących nim zasadach mogą przynieść wymierne efekty i pozwolić na lepsze dostosowanie biznesu do środowiska, w którym działa. Najskuteczniejszą metodą badania specyfiki platform aukcyjnych zdaje się być wykorzystanie usług webowych do pozyskiwania danych i stosowanie technik eksploracji danych w celu wyodrębnienia użytecznej wiedzy.

Zaprezentowane w niniejszym artykule narzędzie AuMiner jest skonstruowane na bazie nowoczesnego paradygmatu architektury zorientowanej na usługi, przy użyciu standaryzowanych protokołów i języków programowania, aby zapewnić jego interoperacyjność, elastyczność działania i możliwość rozszerzania o nowe funkcje w przyszłości. Rozszerzenie narzędzia AuMiner będzie dotyczyło możliwości integrowania danych z wielu portali aukcyjnych. Kolejna wersja będzie zaopatrzona w ontologie pozwalające mapować struktury danych pozyskane z różnych serwisów i przedstawiać je jako jeden spójny model danych nadający się do analizy technikami eksploracji danych. Opisane tutaj rozwiązanie może być zaimplementowane po relatywnie niskim koszcie (przy użyciu narzędzi Open Source) przez osoby prywatne lub przedsiębiorstwa zainteresowane gromadzeniem danych dla



różnych celów analitycznych. Jedyne koszty związane z funkcjonowaniem narzędzia to opłaty dla portalu aukcyjnego uzależnione od liczby wywołań niektórych metod.

## Literatura

1. Chmielarz W., *Próba analizy porównawczej serwisów aukcji internetowych w Polsce*, w: Roczniki Kolegium Analiz Ekonomicznych vol.16/2006.pp.13-26 Wydawnictwo SGH Warszawa 2006.
2. *Collins New compact English dictionary, 2nd edn*, Harper-Collins: Edinburgh, 1992.
3. <http://webapi.allegro.pl/uploader.php> (Listopad 2009).
4. Jureta I., Kolp M., Faulkner S., *An Agent-Oriented Perspective on E-bidding Systems*. In: Vijay Sugumara (ed.), *Intelligent Information Technologies and Applications*, s. 94-118. IGI Publishing, 2007.
5. Ku Y., Chen Y., Chiu C., *A Proposed Data Mining Approach for Internet Auction Fraud Detection*. In: *Intelligence and Security Informatics*. s. 238-243. Springer Berlin, Heidelberg 2007.
6. Mena J., *Investigative Data Mining for Security and Criminal Detection*. Butterworth-Heinemann, Boston 2003.
7. Parente, D.H.(ed.), *Best Practices for Online Procurement Auctions*, IGI Global, USA (2008).
8. Pinker E. A. Seidmann, Y. Vakrat., *Managing online auctions: Current business and research issues*. Management Science, 49(11), s. 1457-1484 2003.
9. Rainardi V., *Building a data warehouse, with examples in SQL server*. Apress United States of America 2008.
10. Pawełoszek-Korek I., *A Framework for Online Auctions Market Research*, w: Proceedings of BIR'2008 the Seventh International Conference on Perspectives in Business Informatics Research. Ed. Stanisław Wrycza. Gdańsk 2008.
11. Haywood A.J., *Online Auctions: User Experience Insights from eBay'*, Chimera Working Paper, Colchester: University of Essex 2006.

---

**PRACTICAL ASPECTS OF ACQUIRING DATA FROM AUCTIONS PORTALS  
FOR KNOWLEDGE EXTRACTION**

**Summary**

Online auctions as a primary or secondary business model are increasingly growing in popularity. The problem is worth analyzing and can be viewed in multiple perspectives. This paper introduces the idea of automated web services-based tool - AuMiner for collecting data from online auction portals. AuMiner is the result of study and work conducted by the author. A framework presented in hereby paper may be implemented at relatively low cost (using open source tools) by organizations or individuals interested in harvesting data from auction services for various analytical purposes like: knowledge discovery in databases for business intelligence or marketing..

*Translated by Ilona Pawełoszek-Korek*