

Dominik Filipiak, Agata Filipowska

Uniwersytet Ekonomiczny w Poznaniu, Wydział Informatyki i Gospodarki Elektronicznej, Katedra Informatyki Ekonomicznej

Autor do korespondencji: Dominik Filipiak, dominik.filipiak@ue.poznan.pl

WIKIPEDIA, DBPEDIA A POLSKI RYNEK SZTUKI: WZBOGACANIE DANYCH ORAZ ANALIZA CZĘSTOŚCI WYSTĘPOWANIA POJĘĆ W OPISACH OBRAZÓW I ARTYSTÓW

Streszczenie: Wikilinki to wewnętrzne hiperłącza w Wikipedii, popularnej internetowej encyklopedii. Unikatowy identyfikator artykułu jest ukryty pod tzw. formą powierzchniową, która jest gramatycznym dopasowaniem danego hasła do kontekstu, w jakim występuje. Tym samym każde hasło może mieć wiele form powierzchniowych.

Wykorzystując to, można w sposób automatyczny wykrywać pojęcia (hasła w Wikipedii) w opisach dzieł i artystów udostępnianych przez domy aukcyjne, wychodząc od wyszukiwania pojawiających się w nich form powierzchniowych. W połączeniu z tradycyjnymi ilościowymi metodami badania rynku sztuki wykorzystującymi regresję liniową podejście to pozwala na wykrycie m.in. popularnych motywów i w konsekwencji zbadanie ich istotności statystycznej oraz wpływu na cenę sprzedaży poszczególnych dzieł.

Słowa kluczowe: rynek sztuki, Wikipedia, regresja hedoniczna.

Klasyfikacja JEL: Z11.

WIKIPEDIA, DBPEDIA AND THE POLISH ART MARKET: DATA ENRICHMENT AND ANALYSIS OF CONCEPTS' FREQUENCY IN THE DESCRIPTIONS OF PAINTINGS AND AUTHORS

Abstract: Wikilinks are internal hyperlinks on Wikipedia, a popular Internet encyclopaedia. A unique article identifier is hidden behind so-called surface form, which

is a grammatical match of a given term accordingly to the context in which it occurs. Therefore, a given term can have multiple surface forms.

Due to this fact it is possible to discover concepts (entries on Wikipedia) automatically in artists' and lots' descriptions on the pages of auction houses by searching for surface forms. Combining these data with traditional quantitative methods of art market research, such as linear regression, it is possible to discover, inter alia, popular motives and explore their statistical significance and their impact on hammer price.

Keywords: art market, Wikipedia, hedonic regression.

Wstęp

Traktowanie dzieła sztuki jako inwestycji nie jest nowym obszarem prac badawczych. Niemniej jednak nie cieszył się on zainteresowaniem naukowców przez długi czas – pierwsze badania zostały przeprowadzone dopiero kilka dekad temu. Koncentrowały się one głównie na opisie trendów, nakreślaniu kształtu rynku oraz temacie wyceny dzieła sztuki. Szczególną uwagę poświęcono budowaniu indeksów, których zadaniem byłoby umożliwienie przejrzystego porównania rynku sztuki z innymi źródłami inwestycji. Budowa takich indeksów wykorzystuje dane rynkowe. W szczególności indeksy hedoniczne budowane są za pomocą dostępnych informacji o dziełach sztuki jako przedmiocie obrotu rynkowego. Tym samym jakość tych danych wpływa na jakość budowanych indeksów, a jej poprawa przekłada się bezpośrednio na jakość rezultatów obliczeń.

W przeciwieństwie do publikacji poświęconych badaniu rynku sztuki w ujęciu ilościowym autorzy w niniejszym artykule obrali sobie za cel poprawę jakości danych o dziełach sztuki z wykorzystaniem danych otwartych (Linked Open Data, dalej: LOD). DBpedia, internetowe źródło danych będące jednocześnie modelowym przykładem LOD, została użyta jako podstawa do wzbogacenia opisu dzieł sztuki. Jest ona wynikiem starań o ustrukturalizowanie wiedzy zgromadzonej w Wikipedii, popularnej internetowej encyklopedii. Ze względu na format przechowywania danych DBpedia umożliwia odpytywanie o zdefiniowane cechy bytów w sposób automatyczny, w przeciwieństwie do Wikipedii. Z jej wykorzystaniem można uzupełnić „luki” w informacjach o rynku sztuki i tym samym poprawić rezultaty w dalszych badaniach, takich jak np. analiza regresji oraz budowanie indeksów.

Artykuł składa się z następujących części. W części pierwszej dokonano przeglądu literatury związanej z przedmiotem badań. Następnie opisano zbiór

danych, zarówno podstawowych, jak i tych używanych do podnoszenia jakości przez wzbogacenie danych pierwotnych. Kolejno omówiono wyniki przeprowadzonych testów empirycznych. Ostatni punkt stanowi krótkie podsumowanie oraz nakreślenie potencjalnych przyszłych kierunków prac badawczych.

1. Przegląd literatury

Ginsburgh, Mei i Moses [2006] w swojej pracy określili zastosowanie indeksów na rynku sztuki jako:

- wskazywanie trendów oraz umożliwienie porównywania inwestycji w sztukę z innymi formami inwestycji (takimi jak np. akcje oraz obligacje),
- pomiar zmienności rynku sztuki oraz jego porównanie z innymi rynkami,
- badanie społecznych oraz ekonomicznych determinant wpływających na rynek sztuki,
- wycenę ogólnej wartości dzieł sztuki.

Podstawową metodą używaną do tworzenia indeksów na rynku sztuki jest wykorzystanie regresji liniowej. Niemniej jednak podejścia do tego tematu można podzielić na dwie grupy [Filipiak i Filipowska 2016]. Pierwszą z nich stanowią indeksy oparte na powtórnej sprzedaży (*repeated-sales regression*). Dotyczą one porównania cen dzieł sprzedanych co najmniej dwukrotnie w rozpatrywanym okresie. Indeksy są budowane na podstawie relacji cen sprzedaży dla tych samych dzieł. Tego typu indeks w założeniach generuje precyzyjne wyniki, ale ze względu na wymóg posiadania w analizowanym zbiorze danych obserwacji z przynajmniej dwukrotną sprzedażą liczba obserwacji możliwych do rozpatrzenia jest bardzo zawężona. W szczególności problem ten dotyka stosunkowo młodych rynków dla dzieł sztuki, takich jak Polska. Do najbardziej znanych indeksów zbudowanych przy użyciu tej metody należy Mei & Moses Fine Art Index.

Druga metoda zakłada budowanie tzw. indeksów hedonicznych [Witkowska i Kompa 2015]. Podobnie jak w poprzedniej metodzie indeks konstruowany jest za pomocą regresji liniowej. Tym razem jednak brane są pod uwagę wszystkie sprzedane dzieła, niekoniecznie dwukrotnie. W swoich założeniach metoda ta interpretuje zmienną objaśnianą (dzieło sztuki) jako zbiór odpowiednio przygotowanych predyktorów ilościowych oraz jakościowych wraz z rokiem jej sprzedaży. Indeks jest konstruowany dzięki wykorzystaniu współczynników związanych z rozpatrywanymi latami sprzedaży. Metoda ta jest silnie związana zarówno z doбором predyktorów, jak i obserwacji i w konsekwencji wymaga lepiej opisanego oraz zbalansowanego zbioru danych. Jest

to rekompensowane przez szerszy zbiór rozpatrywanych dzieł – zniesiony zostaje warunek podwójnej sprzedaży. Regresja hedoniczna znalazła też zastosowanie m.in. w opisywaniu rynku nieruchomości.

Regresja hedoniczna jest, jak już wspomniano, związana z jakością danych. W środowisku akademickim nie ma powszechnej zgody odnośnie do definicji pojęcia jakości danych. Eppler [2006] w swojej pracy wymienił 70 najważniejszych atrybutów danych, z których wskazał te, które mają największy związek z jakością danych. Wśród 16 najistotniejszych atrybutów jest m.in. kompletność (*comprehensiveness*), dostępność (*accessibility*) oraz odpowiedniość (*applicability*). Kräussl oraz Elsland [2008] analizowali użycie zmiennych objaśniających w modelach hedonicznych. Wśród tych najpopularniejszych wskazali oni m.in. rok sprzedaży, rozmiar, powierzchnię, technikę, miejsce sprzedaży, dom aukcyjny, występowanie sygnatury, wiek autora, średnią cenę sprzedaży dzieł danego autora, liczbę wystaw, na których obecne było dane dzieło, oraz proveniencję. Jest więc oczywiste, że dysponując większym zakresem danych, można budować bardziej skomplikowane i precyzyjne modele. Do oceny modelu tradycyjnie wykorzystywany jest współczynnik determinacji, ale nie jest to jedyna miara.

W ostatnich latach można zaobserwować zwiększone zainteresowanie badaczy pojęciem Linked Open Data. Do najbardziej znanych źródeł danych zaliczanych do tej kategorii należy DBpedia (<http://dbpedia.org>) [Lehmann i in. 2015]. Tworząc ten zasób, wykorzystano część informacji o zdefiniowanej strukturze zawartych w Wikipedii (tzw. infoboksy). W technicznym ujęciu DBpedia stanowi bazę danych trójek zbudowaną z wyekstrahowanych informacji zawartych w infoboksach. Dzięki temu te dane można w wygodny sposób pozyskiwać za pomocą zapytań SPARQL. Podczas gdy wysiłki badaczy skupiają się na metodach opisu oraz publikacji otwartych danych, warto przyrzeć się ich praktycznemu wykorzystaniu. Pośród wielu różnych dziedzin, w których używano takich danych, Filipiak oraz Filipowska [2015] opisują schemat postępowania w przypadku wykorzystania źródeł danych na potrzeby rynku sztuki.

2. Opis dostępnych danych

Użyte w eksperymentach dane dotyczą sprzedaży dzieł sztuki w domu aukcyjnym Desa Unicum, głównym przedstawicielu tego segmentu na polskim rynku (<https://desa.pl/pl/news/91/rekordowy-rok-raport-2015>). Pod uwagę zostały wzięte dzieła sprzedane między 2003 a 2013 rokiem. Dla każdej obserwacji, tj. dzieła wystawionego na sprzedaż, wyróżniono następujące

cechy: autor, cena sprzedaży, dolna estymacja, górna estymacja, opis, medium, rozmiar, data aukcji, kategoria aukcji, tytuł, rok urodzenia oraz śmierci autora. Warto zauważyć, że dane w testowanym zbiorze często były niekompletne. W szczególności tyczy się to ceny sprzedaży, gdyż wynika to ze specyfiki aukcji – nie każde dzieło musi znaleźć nowego nabywcę. Innym przykładem jest brak opisu oraz dat dla mniej znanych autorów i dzieł. Tego typu braki obniżają jakość danych, rozumianą w tym artykule przede wszystkim jako ich kompletność. Stanowi to motywację do jej poprawienia. W tabeli 1 zostały opisane podstawowe charakterystyki ilościowe dla pozyskanego zbioru danych.

Tabela 1. Podstawowe charakterystyki ilościowe opisujące korpus danych pozyskanych z domu aukcyjnego Desa Unicum oraz form powierzchniowych w Wikipedii

Charakterystyka	Liczba
Liczba obserwacji w zbiorze danych	25 837
Liczba unikatowych autorów w zbiorze danych	5 691
Liczba form powierzchniowych	3 2449 566
Liczba unikatowych artykułów wskazywanych przez formy powierzchniowe	5 056 923
Liczba autorów powiązanych dzięki formom powierzchniowym	1 332

Źródło: na podstawie danych ze strony domu aukcyjnego Desa Unicum oraz zrzutu polskiej Wikipedii.

Drugim istotnym źródłem danych w prowadzonych badaniach był słownik mapujący formy powierzchniowe występujące w opisach dzieł sztuki na byty nazwane (*named entities*). W badanym przypadku sprowadza się to do łączenia różnych form gramatycznych (np. „kościelne”, „kościół”) z poszczególnymi artykułami w Wikipedii („Kościół (budynek)”) przy użyciu wyrażeń regularnych. Słownik ten został zbudowany za pomocą kompletnego zrzutu zawartości artykułów polskiej Wikipedii, w którym wikilinki oznaczone są specjalną składnią. W tabeli 1 zostały opisane także podstawowe charakterystyki ilościowe związane z tym zbiorem danych.

Dodatkowym źródłem danych pozwalającym na sprowadzenie danego wyrazu do jego formy podstawowej jest słownik Polimorf, który został opisany szerzej w kolejnej części artykułu.

3. Metodologia prac badawczych

Wzbogacenie danych dostępnych dla pojedynczego sprzedawanego przedmiotu może się odbyć na dwa sposoby. Pierwszy z nich zakłada wykorzystanie

nie informacji o autorze, tj. jego imieniu oraz nazwisku. Drugi sposób jest związany z analizą tekstu opisu dzieła i wykorzystuje techniki przetwarzania języka naturalnego.

Pierwsza metoda ma na celu powiązanie autorów z poszczególnych obserwacji z odpowiadającymi im artykułami w Wikipedii (i tym samym encjami DBpedii). Podstawową trudnością tej metody jest to, że dane osobowe twórcy można zapisać na wiele sposobów i często różnią się one od tych opisujących stosowny artykuł. Jako przykład można podać dzieła Witkacego. W zbiorze danych autor ten reprezentowany jest jako „Witkacy”, natomiast odpowiadający mu wpis w Wikipedii to „Stanisław Ignacy Witkiewicz”. Z punktu widzenia automatycznego przetwarzania stanowi to więc problem. Zastosowanym przez autorów rozwiązaniem jest użycie wikilinków. Są to wewnętrzne hiperłącza Wikipedii. W ich konstrukcji unikatowy identyfikator artykułu jest ukryty pod tzw. formą powierzchniową, która jest gramatycznym dopasowaniem danego hasła do kontekstu, w jakim występuje. Tym samym każde hasło (np. polska conceptualistka „Natalia LL”) może mieć wiele form powierzchniowych wynikających z różnych form gramatycznych, skrótów i kontekstu użycia (np. „Natalii LL”, „Natalia Lach-Lachowicz”). Dzięki temu można w sposób automatyczny wykrywać koncepty (hasła w Wikipedii) w danych udostępnianych przez domy aukcyjne, wychodząc od wyszukiwania w nich form powierzchniowych. W badaniach przygotowano słownik form powierzchniowych na podstawie zrzutu zawartości artykułów polskiej Wikipedii. Prosta konstrukcja słownika (forma powierzchniowa wskazująca na artykuł) powodowała jednak nieco trudności, gdyż niektóre formy powierzchniowe mogą wskazywać na kilka różnych artykułów. Problem wieloznaczności nie jest trywialny i na potrzeby tych badań takie przypadki zostały odrzucone.

Druga metoda polega na analizie tekstu dzieła i częstości występowania pojęć opisujących dzieło w jego opisie. Należy pamiętać, że metoda ta stawia znacznie więcej wyzwań. Precyzując – jest to dokładnie taka sama skala trudności, co przy przetwarzaniu dowolnego tekstu w języku naturalnym. Zadania z tego obszaru zazwyczaj można podzielić na kilka etapów, do których zaliczamy lematyzację oraz rozpoznawanie bytów nazwanych. Przez lematyzację w przeprowadzonych badaniach rozumie się sprowadzenie każdego słowa w opisie dzieła sztuki do kanonicznej formy leksemu, który mu odpowiada (np. „kościół” do „kościół”). Aby sprostać temu zadaniu, potrzebny jest rozbudowany zbiór danych. Zastosowanym słownikiem morfologicznym dla języka polskiego był Polimorf [Woliński i in. 2012]. Powstał on jako połączenie rezultatów prac nad słownikami Morfeusz SGJP oraz Morfologik

i jest najbardziej rozbudowanym narzędziem tego typu dla języka polskiego w chwili pisania tego artykułu.

Przygotowano także zbiór słów wchodzących w skład specjalnego słownika tzw. stop listy (*stop-words*). Zbiór ten jest wykorzystywany do usuwania wymienionych w nim słów z przetwarzanych opisów dzieł. Zasadniczo składa się on z dwóch części: zbioru słów o małym znaczeniu, który jest tradycyjnie używany w stop listach (chodzi m.in. o spójniki czy też popularne słowa, takie jak „jest”) oraz zbioru słów o niskiej wartości dyskryminacyjnej w przetwarzanym zbiorze danych związanych z domeną zbioru danych. W badanym przypadku słowa te to m.in. „estymacja” oraz „odwrocie” – powtarzają się one w większości opisów i nie są związane bezpośrednio z charakterystyką dzieła sztuki.

Po podzieleniu opisu na terminy oraz po ich lematyzacji kolejnym krokiem tej metody jest rozpoznawanie bytów nazwanych. Zasada działania jest tu identyczna jak dla pierwszej metody, tj. wykorzystuje się słownik form powierzchniowych.

4. Wyniki eksperymentu

Do przeprowadzenia eksperymentu użyto technologii Apache Spark (<http://spark.apache.org>), która umożliwia szybkie i wygodne przeprowadzenie równoległych obliczeń, w szczególności w pracy z plikami tekstowymi [Zaharia i in. 2010]. Za pomocą języka SPARQL [Hartig, Bizer i Freytag 2009] napisano zapytania, które pozwalały powiązać łańcuchy znaków reprezentujące autorów (artystów) z bytami nazwanymi reprezentującymi ich w DBpedii. W pierwszym eksperymencie zbadano wykorzystanie polskiej DBpedii do uzyskania informacji o autorach. Za pomocą języka SPARQL dla każdego z autorów w sposób automatyczny wykonano zapytanie o jego datę urodzenia (własność *dbpedia-owl:birthDate*) oraz śmierci (własność *dbpedia-owl:deathDate*), styl (własność *prop-pl:style*) oraz narodowość (własność *prop-pl:narodowość*). Wyniki dla znalezionych encji w ujęciu ogólnym oraz tylko dla unikatowych autorów przedstawiono w tabeli 2.

O ile data urodzenia oraz śmierci autora są jednoznacznymi wartościami charakteryzującymi się większą lub mniejszą precyzją, o tyle w przypadku narodowości (tabela 3) oraz stylu (tabela 4) sytuacja jest bardziej skomplikowana. W związku z tym, że wykorzystano polską DBpedię oraz zbiór danych pochodzący z polskiego domu aukcyjnego, właśnie ta narodowość dominuje w statystykach. Rezultaty wskazują na to, że pomimo wykorzystania w założeniu ustrukturyzowanego źródła danych zwrócone wartości nie wskazują

Tabela 2. Wynik informacji znalezionych w polskiej DBpedii na podstawie danych osobowych autorów

Atrybut	Własność RDF	Liczba oraz procent znalezionych encji dla całego zbioru danych	Liczba oraz procent znalezionych encji dla unikatowych autorów danych
Data urodzenia	<i>dbpedia-owl:birthDate</i>	4145 (16,04)	386 (6,68)
Data śmierci	<i>dbpedia-owl:deathDate</i>	3556 (13,76)	286 (5,02)
Styl	<i>prop-pl:styl</i>	1942 (7,52)	119 (2,09)
Narodowość	<i>prop-pl:narodowość</i>	2850 (11,03)	210 (3,69)

Źródło: na podstawie badań empirycznych.

jednoznacznie na pochodzenie autorów. „Polska”, „Polak” oraz „Polka” co prawda intuicyjnie odnoszą się do tej samej narodowości, ale z punktu widzenia zastosowania wyników w analizie regresji należy poddać je dodatkowej obróbce i sprowadzeniu do jednego pojęcia. Rozwiązaniem problemu może być użycie rozmytego dopasowywania ciągów znaków.

Tabela 3. Narodowość – najpopularniejsze wartości

Atrybut	Liczba znalezionych encji
„polska”	1181
< http://pl.dbpedia.org/resource/Polacy >	1099
„Polak”	217
< http://pl.dbpedia.org/resource/Lemkowie >	117
< http://pl.dbpedia.org/resource/Polska >	72
„Polka”	44
„Polska”	38
„amerykańska”	16

Źródło: na podstawie badań empirycznych.

Sytuacja wygląda lepiej w przypadku stylów reprezentowanych przez autorów (tabela 4). Co prawda problem znany z przypadku narodowości również występuje, jednak w o wiele mniejszej skali (przypadek symbolizmu). Realizm, symbolizm oraz impresjonizm są dominującymi kierunkami wśród sprzedawanych dzieł sztuki w badanym zbiorze danych.

Rezultaty rozpoznawania bytów nazwanych za pomocą form powierzchniowych (tabela 5) sugerują, że problem jest złożony – popularne koncepty, takie jak „Malarstwo” lub „Galeria sztuki”, przewijają się w większości opisów dzieł, ale nie wnoszą nic z punktu widzenia ilościowych badań rynku sztuki.

Tabela 4. Styl – najpopularniejsze wartości

Atrybut	Liczba znalezionych encji
< http://pl.dbpedia.org/resource/Realizm_(malarstwo) >	259
< http://pl.dbpedia.org/resource/Symbolizm >	213
< http://pl.dbpedia.org/resource/Impresjonizm >	183
< http://pl.dbpedia.org/resource/Prymitywizm_(malarstwo) >	117
< http://pl.dbpedia.org/resource/Modernizm_(sztuka) >	116
„symbolizm”	111
< http://pl.dbpedia.org/resource/Ekspresjonizm_(sztuka) >	108

Źródło: Na podstawie badań empirycznych.

O wiele ciekawsze może być natomiast badanie miejsc (m.in. „Warszawa”, „Łódź”, „Paryż”) czy też bytów związanych z motywami obecnymi w poszczególnych dziełach („Kobieta” występowała 701 razy, „Dom” – 625, „Kościół (budynek)” – 583). Niemniej jednak podstawowym problemem pozostaje usunięcie ze zbioru bytów o niskiej wartości dyskryminacyjnej, co sprowadzi się do znacznego powiększenia stop listy. Ponieważ proces ten może być niezwykle czasochłonny, z pomocą mogą przyjść algorytmy, takie jak np. TF-IDF (*Term Frequency–Inversed Document Frequency*), które biorą pod uwagę częstotliwość występowania termów w dokumentach (tu: opisach) i jednocześnie sprawdzają liczbę dokumentów w całym zbiorze, w których ten term występuje.

Tabela 5. Najczęściej występujące pojęcia w opisie dzieł sztuki

Pojęcie	Liczba wystąpień pojęcia
Malarstwo	17559
Galeria sztuki	9966
Artysta	9242
Rok	7972
Warszawa	7176
Sztuka	6949
Polska	6778
Pracownia	4931
Grafika	4622
Piękno	4556
Datowanie	4299

Źródło: na podstawie badań empirycznych.

Podsumowanie

W artykule zaprezentowano dwie metody wzbogacania danych dla rynku sztuki celem polepszenia jakości budowanych indeksów. Pierwsza jest związana z uzupełnieniem danych twórcy, a druga z poszerzeniem opisu dzieła. W połączeniu z tradycyjnymi ilościowymi metodami badania rynku sztuki wykorzystującymi regresję liniową poprzez włączenie tych danych do analizowanego zbioru może to pozwolić na wykrycie m.in. popularnych motywów i w konsekwencji zbadanie ich istotności statystycznej oraz wpływu na cenę sprzedaży poszczególnych dzieł.

O ile pierwsza metoda daje zadowalające rezultaty i pozwala w pewnym stopniu uzupełnić braki w danych, o tyle druga metoda z pewnością wymaga dalszych prac. Zastosowanie bardziej wyszukanych metod rozpoznawania bytów nazwanych bez wątplenia przyczyni się do poprawy wyników przedstawionej metody. Oczywistym kierunkiem badawczym pod względem empirycznym wydaje się też rozszerzenie zbioru danych do co najmniej czterech najpopularniejszych polskich domów aukcyjnych (Agra Art, Desa Unicum, Polswiss Art, Rempex). Finałowym etapem badań powinno być zbudowanie polskiego indeksu rynku sztuki, już na podstawie wzbogaconych danych.

Bibliografia

- Eppler, M.J., 2006, *Managing Information Quality*, Springer, Berlin, Heidelberg.
- Filipiak, D., Filipowska, A., 2015, *DBpedia in the Art Market*, Business Information Systems Workshops, vol. 228, s. 321–331.
- Filipiak, D., Filipowska, A., 2016, *Towards Data-oriented Analysis of the Art Market*, Financial Internet Quarterly "e-Finanse", vol. 12, no. 1, s. 21–31.
- Ginsburgh, V., Mei, J., Moses, M., 2006, *The Computation of Prices Indices*, in: Throsby, C.D., Ginsburgh, V. (eds.), *Handbook of the Economics of Art and Culture*, Elsevier, North Holland, s. 947–979.
- Hartig, O., Bizer, C., Freytag, J.C., 2009, *Executing SPARQL Queries over the Web of Linked Data*, in: *8th International Semantic Web Conference*, Chantilly, VA, Springer, Berlin Heidelberg, s. 293–309.
- Kräussl, R., van Elsland, N., 2008, *Constructing the True Art Market Index: A Novel 2-step Hedonic Approach and Its Application to the German Art Market*, CFS Working Paper Series, Center for Financial Studies, Frankfurt, no. 11.

- Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., Bizer, C., 2015, *DBpedia–A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia*, *Semantic Web Journal*, vol. 6, s. 167–195.
- Witkowska, D., Kompa, K., 2015, *Constructing Hedonic Art Price Indexes for the Polish Painting Market Using Direct and Indirect Approaches*, *AESTIMO*, *The IEB International Journal of Finance*, vol. 10, s. 2–25.
- Woliński, M., Miłkowski, M., Ogrodniczuk, M., Przepiórkowski, A., Szalkiewicz, Ł., 2012, *PoliMorf: A (not so) New Open Morphological Dictionary for Polish*, in: *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, Istanbul, s. 860–864.
- Zaharia, M., Chowdhury, M., Franklin, M.J., Shenker, S., Stoica, I., 2010, *Spark: Cluster Computing with Working Sets*, in: *Proceedings of the 2nd USENIX Conference on Hot Topics in Cloud Computing*, USENIX Association, Berkeley, CA, s. 10–10.