

Politechnika Poznańska
Wydział Informatyki
Instytut Informatyki

Praca dyplomowa magisterska
**Opracowanie i implementacja systemu
analizującego dane bibliometryczne**
Mikołaj Ignaszak

Promotor:
Prof. Jacek Błazewicz

Opiekun naukowy:
Dr inż. Szymon Wąsik

Poznań 2016 r.

Miejsce na kartę pracy

Spis treści

1.	Wstęp	7
2.	Zbieranie danych bibliograficznych	9
2.1	Ocena dorobku naukowego	9
2.2	Podstawy bibliometrii	9
2.3	Czasopisma punktowane	10
2.4	Definicje typów publikacji naukowych	11
3.	Źródła informacji o cytowaniach	13
3.1	Google Scholar	13
3.2	Web Of Science	16
3.3	Scopus	19
3.4	Microsoft Academic	22
3.5	Scholar+	23
3.6	Podsumowanie	26
4.	Projekt wzbogacenia funkcjonalności baz bibliometrycznych	29
4.1	System Scholar+	29
4.2	Analiza możliwości wzbogacenia istniejących serwisów	31
4.3	Opis założeń Web Of Science +	33
5.	Etapy projektu	37
5.1	Wymagania funkcjonalne	37
5.2	Wykorzystane technologie	37
5.3	Projekt	38
5.4	Implementacja	39
5.5	Efekt końcowy	44
6.	Zakończenie	47
7.	Dodatki	49
	Dodatek A. Spis zawartości płyty CD dołączonej do pracy	49
	Dodatek B. Instrukcja instalacji i konfiguracji systemu Web of Science+	51
	Dodatek C. Kody źródłowe wybranych funkcji aplikacji Web Of Science +	53
8.	Bibliografia	61
9.	Spis ilustracji i tabel	63

1. Wstęp

Ważną kwestią w karierze pracownika naukowego jest poprawna i obiektywna ocena jego dorobku. Każdy naukowiec potrzebuje informacji, jak bardzo i w jaki sposób jego publikacje odznaczają się w środowisku naukowym. Natomiast jednostka organizacyjna (na przykład wydział) potrzebuje danych poprawnie oddających wpływ prac badaczy na inne wydawnictwa w celu odpowiedniego i sprawiedliwego oceniania pracowników. Statystyki autora są również ważne przy awansach oraz przy otwieraniu przewodu doktorskiego i habilitacyjnego.

Zestaw narzędzi matematycznych pomagających w ocenie bibliografii naukowców nazywany jest bibliometrią. Posługuje się ona zestawem wskaźników ilościowych oceniających publikacje oraz klasyfikujących czasopisma i konferencje naukowe. Jest to nietrywialne, ponieważ cytowania różnią się wagą w zależności od miejsca publikacji. Dlatego wartość publikacji będzie różna w zależności od wydawnictwa jak i źródła cytowań. Lista najważniejszych czasopism i konferencji jest publikowana przez Ministerstwo Nauki i Szkolnictwa Wyższego.

Istnieje wiele narzędzi służących do zbierania i przeglądania danych bibliograficznych oraz obliczających statystyki ułatwiające ocenę pracy i dorobku naukowca. Różnią się one wielkością i jakością bazy, funkcjonalnością czy typem licencji. Różnorodność dostępnych systemów pozwala użytkownikowi wybrać i korzystać z rozwiązania, które mu odpowiada. Niestety duże zróżnicowanie serwisów powoduje, że żaden z nich nie jest idealny i nie spełnia w pełni wymagań poszczególnych użytkowników.

1.1 Cel i zakres pracy

Celem pracy jest stworzenie serwisu, który będzie rozszerzał jedno z istniejących narzędzi bibliograficznych. Będzie niwelować niedogodności występujące w istniejących portalach oraz wprowadzać nową funkcjonalność. Przed rozpoczęciem prac nad projektem należy dokładnie przeanalizować dostępne rozwiązania pod kątem posiadanych wad oraz możliwości ich rozszerzanie. Wymagane jest stworzenie zestawienia baz publikacji i cytowań pod względem ich jakości i wielkości.

Po wybraniu najlepszej podstawy do realizacji pracy należy spisać wymagania funkcjonalne do aplikacji. Powinna ona w pełni spełniać potrzeby użytkowników oraz być napisana tak, aby umożliwiać jej łatwy rozwój w przyszłości.

Zakres pracy obejmuje:

1. Analizę dostępnych narzędzi bibliograficznych.
2. Wybór jednego z nich jako podstawy projektu.
3. Realizacja projektu (spis wymagań, wstępny projekt aplikacji, implementacja).

2. Zbieranie danych bibliograficznych

Dokumentacja historii publikacji naukowych jest istotna w kontekście oceny dorobku naukowego. Pomaga w tym bibliometria. Po raz pierwszy termin ten zastosował Alan Pritchard w pracy *Statistical Bibliography or Bibliometrics*. Oznacza on zastosowanie matematycznych i statystycznych metod do badania książek i innych środków komunikacji [3]. Według Głównego Urzędu Statystycznego głównym źródłem danych bibliometrycznych jest zespół baz danych: Science Citation Index (SCI), Social Science Citation Index (SSCI) oraz Arts & Humanities Citation Index (A&HCI), opracowanych przez Instytut Informacji Naukowej (Institute for Scientific Information - ISI, Filadelfia, USA) [5].

2.1 Ocena dorobku naukowego

Ocena pracy pracownika naukowego odbywa się na dwóch poziomach: lokalnym (instytut, wydział, uczelnia) oraz globalnym. Na poziomie lokalnym dana jednostka przydziela każdemu zatrudnionemu naukowcowi punkty. Na ich podstawie między innymi przyznaje stypendium lub umożliwia otwarcie przewodu doktorskiego i habilitacyjnego. Kryteria lokalne są ustalane indywidualnie przez daną komórkę. Na poziomie globalnym punkty otrzymuje jednostka dydaktyczna, na podstawie dokonań i aktywności wszystkich pracowników. Są one przydzielane według jasno ustalonych zasad dla wszystkich jednostek w danej dziedzinie naukowej. Punkty można otrzymać między innymi za:

- publikowanie monografii, rozdziałów, artykułów w czasopismach punktowanych znajdujących się w wykazie Ministerstwa Nauki i Szkolnictwa Wyższego [11],
- organizację konferencji,
- opatentowanie wynalazku,
- wydawanie czasopisma naukowego [4].

System punktacji na poziomie lokalnym zazwyczaj nie jest równoznaczny z systemem globalnym, choć często jest nim inspirowany, ponieważ jednostce zależy na tym, aby jej pracownicy pracowali na dobry, globalny wynik jednostki

2.2 Podstawy bibliometrii

Bibliometria posługuje się wskaźnikami ilościowymi. Podstawowym parametrem jest Impact Factor (inaczej wskaźnik wpływu), który służy do oceny czasopism. Określa liczbę cytowań uzyskanych przez artykuły publikowane w tym wydawnictwie w okresie dwóch i pięciu lat, podzieloną przez ogólną liczbę artykułów zamieszczonych w tym czasopiśmie w danym okresie [3]. Budowany jest na podstawie bazy publikacji naukowych. Jedną z nich jest tzw. baza filadelfijska tworzona przez Institute of Scientific Information (ISI) (popularnie zwany instytutem filadelfijskim). Indeksowanie wszystkich czasopism naukowych jest

niemożliwe technicznie, dlatego instytut wybiera wydawnictwa według prawa Bradforda. Mówi ono, że w każdej dziedzinie naukowej istnieje określona grupa pism (od 5 do 30), w których publikowane jest 99% naprawdę istotnych artykułów. Lista filadelfijska zawiera około 8400 pozycji. Spis większości czasopism z listy filadelfijskiej wraz z IF jest publikowany corocznie w raporcie Journal Citation Reports (JCR) przez komercyjną firmę Thomson Reuters [19]. Spis ten choć inspirowany listą filadelfijską nie pokrywa się z nią w pełni, gdyż jego skład zależy od uznania prywatnego wydawcy.

Najprostszym sposobem oceny pracy i dorobku naukowego danego pracownika jest zsumowanie liczby cytowań jego publikacji w wydawnictwach z wybranej listy. Przydatne jest też wyliczenie średniej liczby cytowań na publikację. Niestety oba te rozwiązania mają swoje wady, między innymi nie pokazują prawdziwej wagi i jakości publikacji autora.

Popularnym wskaźnikiem przedstawiającym dorobek naukowy i jakość cytowań prac autora jest indeks Hirscha. Powołując się na doktor Helenę Dryzek możemy go zdefiniować następująco:

„Współczynnik h dla danego autora jest to liczba publikacji cytowanych $\geq h$ razy. Na przykład współczynnik $h = 10$ oznacza, że autor ma 10 publikacji cytowanych co najmniej 10 razy. Wielkość h zależy więc od dwóch czynników: liczby publikacji i ich popularności. Wzrost współczynnika h można osiągnąć publikując prace, które znajdą znaczny oddźwięk.” [20]

Wskaźnik ten nagradza naukowców, którzy publikują mniejszą liczbę prac, względem tych, których historia badań jest szeroka, lecz rozproszona. Premiuje także autorów, których publikacje są stałe i często cytowane, gdyż z biegiem czasu indeks h dla nich wzrasta. Stosowanie tego wskaźnika ma sens, gdy porównujemy naukowców pracujących w tej samej dziedzinie [18]. Do obliczania indeksu Hirscha służy wiele narzędzi. Niektórymi z nich są:

- bazy ISI Web of Knowledge,
- bazy SCOPUS,
- programu „Publish or Perish”,
- Google Scholar Citations [6].

2.3 Czasopisma punktowane

Wykaz czasopism punktowanych jest publikowany przez Ministerstwo Nauki i Szkolnictwa Wyższego. Za publikowanie w wydawnictwach z wyżej wymienionej listy są przyznawane punkty dla jednostki organizacyjnej, do której należy autor. Zestawienie jest podzielone na 3 grupy:

- lista A: są to czasopisma występujące na liście Journal Citation Reports (JCR), posiadające Impact Factor. Za publikację w pismach z tej listy można dostać maksymalnie 50 punktów,

- lista B: jest to lista czasopism nieposiadających Impact Factora. Maksymalna liczba punktów dla publikacji w wydawnictwie z tej listy to 15 punktów,
- lista C: jest to lista wydawnictw dla publikacji humanistycznych uwzględnionych w bazie European Reference Index for the Humanities (ERIH). Maksymalna liczba punktów to 25 [7].

2.4 Definicje typów publikacji naukowych

Autorzy dostają różną ilość punktów w zależności od typu publikacji. Warto w tym momencie przytoczyć definicje z rozporządzenia Ministra Nauki i Szkolnictwa Wyższego z dnia 27 października 2015 r. w sprawie kryteriów i trybu przyznawania kategorii naukowej jednostkom naukowym. Rozróżnia się w nim następujące typy publikacji:

- artykuł naukowy

§8.2 Przez publikację rozumie się recenzowany artykuł naukowy zamieszczony w czasopiśmie naukowym, prezentujący wyniki badań naukowych lub prac rozwojowych o charakterze empirycznym, teoretycznym, technicznym lub analitycznym, przedstawiający metodykę badań naukowych lub prac rozwojowych, przebieg procesu badawczego i jego wyniki, wnioski – z podaniem cytowanej literatury (bibliografię). Do artykułów naukowych zalicza się także opublikowane w czasopismach naukowych recenzowane opracowania o charakterze monograficznym, polemicznym lub przeglądowym oraz glosy lub komentarze prawnicze.

- monografia

§8.3 . Monografie naukowe, w tym edycje naukowe tekstów źródłowych i artystycznych, atlasy i mapy, tematyczne encyklopedie i leksykony, komentarze do ustaw, słowniki biograficzne i bibliograficzne, bibliografie oraz katalogi zabytków, zalicza się do osiągnięć naukowych i twórczych jednostki naukowej, jeżeli spełniają łącznie następujące warunki:

- *stanowią spójne tematycznie opracowania naukowe,*
- *przedstawiają określone zagadnienie w sposób oryginalny i twórczy,*
- *poddane były procedurze recenzji wydawniczych,*
- *są opatrzone właściwym aparatem naukowym (bibliografia lub przypisy), z wyłączeniem map,*
- *posiadają objętość co najmniej 6 arkuszy wydawniczych lub są mapami odpowiadającymi tej objętości tekstu,*
- *są opublikowane jako książki lub odrębne tomy (z wyłączeniem map), których egzemplarze obowiązkowe zostały przekazane uprawnionym bibliotekom, zgodnie z art. 3 ustawy z dnia 7 listopada 1996 r. o obowiązkowych egzemplarzach*

bibliotecznych (Dz. U. Nr 152, poz. 722, z późn. zm.), lub są opublikowane w formie elektronicznej w Internecie,

- *posiadają nadany numer ISBN lub ISMN lub identyfikator DOI (Digital Object Identifier – cyfrowy identyfikator dokumentu elektronicznego).*

§8.6 Do monografii nie zalicza się skryptów akademickich i podręczników akademickich, jeżeli nie spełniają wszystkich warunków określonych w ust. 3, monograficznych artykułów opublikowanych w czasopiśmie, powieści, zbiorów poezji, zbiorów opowiadań i reportaży, pamiętników i dzienników oraz wznowień monografii naukowych.

- *monografia wybitna*

§8.4 Podstawę do uznania monografii naukowej za dzieło „wybitne” stanowi prestiżowa nagroda, przyznana w okresie objętym oceną: nagroda Prezesa Rady Ministrów, ministra, właściwego Wydziału Polskiej Akademii Nauk, Komitetu Naukowego Polskiej Akademii Nauk, Fundacji na Rzecz Nauki Polskiej, zagranicznego towarzystwa naukowego, organizacji międzynarodowej lub ogólnopolskiego towarzystwa naukowego o szczególnym prestiżu.

- *rozdział w monografii*

§8.7 Do osiągnięć naukowych i twórczych jednostki naukowej zalicza się rozdział w monografii naukowej stanowiący opracowanie naukowe o objętości co najmniej pół arkusza wydawniczego lub odpowiadające tej objętości tekstu odrębnie publikowane mapy oraz hasła opublikowane w wydawnictwach encyklopedycznych i słownikowych o objętości co najmniej 0,25 arkusza wydawniczego, spełniających wymagania określone w ust. 3

3. Źródła informacji o cytowaniach

Istota zbierania danych bibliograficznych i generowania statystyk na ich podstawie w kontekście dorobku naukowego niesie za sobą potrzebę odpowiedniego ich przechowywania i łatwego dostępu dla zainteresowanych (pracowników naukowych, jednostek). Istnieje kilka serwisów, które udostępniają informacje na temat publikacji autora oraz ilości i jakości ich cytowań. Strony te różnią się przede wszystkim wielkością bazy danych, ale również prezentowanymi informacjami czy dostępnymi narzędziami. W tym rozdziale zostaną opisane najpopularniejsze z nich.

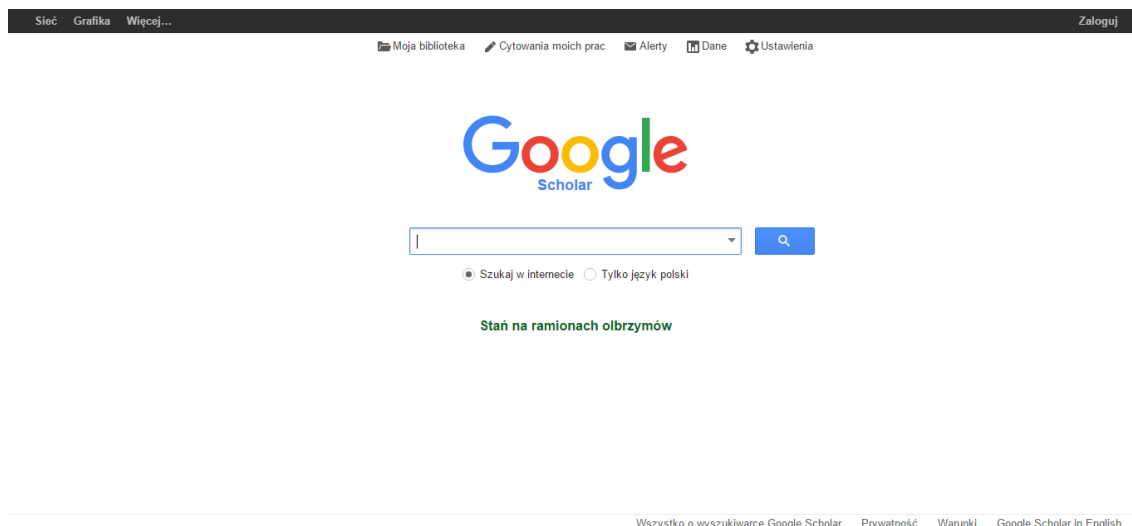
3.1 Google Scholar

Najbardziej popularnym ogólnodostępnym i darmowym narzędziem do wyszukiwania publikacji naukowych i ich cytowań jest założony przez amerykańską firmę Google portal Google Scholar. Wyszukiwarka wystartowała w 2004 roku. Strona scholar.google.pl określa charakterystykę usługi następująco: *„Google Scholar provides a simple way to broadly search for scholarly literature. From one place, you can search across many disciplines and sources: articles, theses, books, abstracts and court opinions, from academic publishers, professional societies, online repositories, universities and other web sites. Google Scholar helps you find relevant work across the world of scholarly research.”* Funkcjonalność Google Scholar:

1. Przeszukiwanie literatury naukowej.
2. Badanie powiązanych prac, cytowań, autorów oraz publikacji.
3. Umieszczanie kompletnych dokumentów za pomocą biblioteki lub Internetu.
4. Dostęp do najnowszych osiągnięć w każdej dziedzinie naukowej.
5. Tworzenie publicznego profilu autora [22].

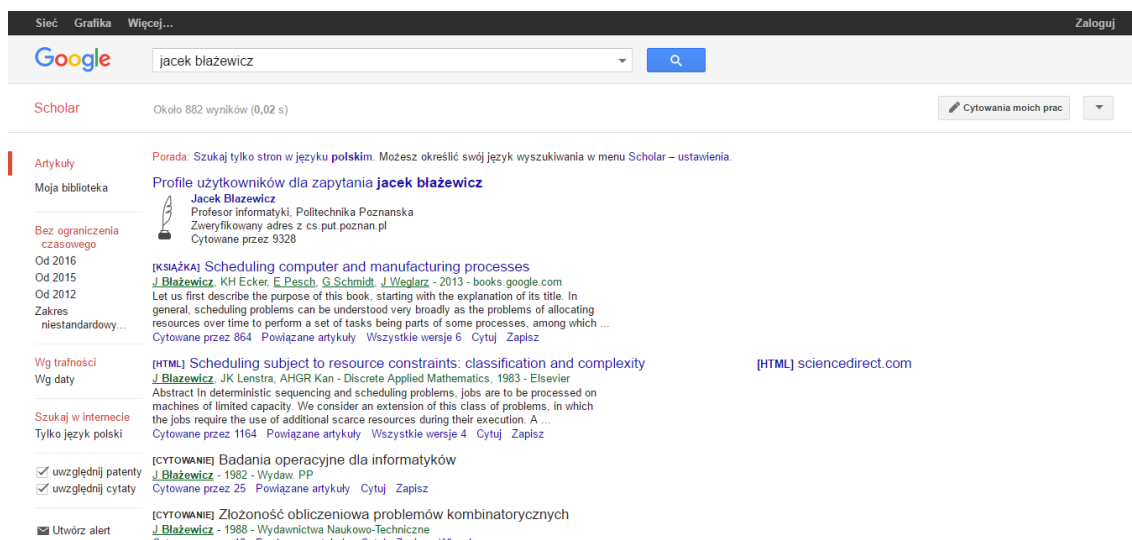
Google nie publikuje informacji o wielkości bazy danych w Google Scholar. Możemy jednak odnieść się do badań, które informują że znajduje się tam około 160 milionów dokumentów (stan na 2014 rok) [1].

Google Scholar stara się klasyfikować dokumenty tak samo jak robią to badacze, oceniając tekst, miejsce publikacji, autorów oraz częstotliwość cytowań w literaturze naukowej w ostatnim czasie [22].



Rysunek 1 Główna strona Google Scholar

Po wejściu na stronę scholar.google.pl wyświetla się wyszukiwarka. Po wpisaniu szukanej frazy (nazwisko autora, nazwa pracy) następuje przekierowanie do listy znalezionych elementów.




Rysunek 2 Lista wyszukanych elementów

Na liście znajdują się znalezione profile, publikacje jak i cytowania. Istnieje możliwość uściślenia wyszukiwania według dat publikacji.

Sieć Grafika Więcej...

Zaloguj



Jacek Blazewicz

Profesor informatyki, Politechnika Poznańska
 informatyka, teoria szeregowania zadań, bioinformatyka
 Zweryfikowany adres z cs.put.poznan.pl - Strona główna

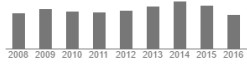
Obserwuj

Tytuł	1-20	Cytowane przez	Rok
Scheduling subject to resource constraints: classification and complexity	J Blazewicz, JK Lenstra, AHGR Kan Discrete Applied Mathematics 5 (1), 11-24	1164	1983
Scheduling computer and manufacturing processes	J Blazewicz, KH Ecker, E Pesch, G Schmidt, J Weglarz Springer Science & Business Media	864	2013
The job shop scheduling problem: Conventional and new solution techniques	J Blazewicz, W Domschke, E Pesch European journal of operational research 93 (1), 1-33	517	1996
Scheduling in computer and manufacturing systems	J Blazewicz, KH Ecker, G Schmidt, J Weglarz Springer Science & Business Media	499	2012
Sequencing of parts and robot moves in a robotic cell	SP Sethi, C Srisankarajah, G Sorger, J Blazewicz, W Kubiak International Journal of Flexible Manufacturing Systems 4 (3-4), 331-358	295	1992

Google Scholar

Indeksy cytowań

	Wszystkie	Od 2011
Cytowania	9328	3406
h-indeks	42	24
i10-indeks	136	75



Współautorzy

Wyświetl wszystkich...

Jan Weglarz
 Erwin Pesch
 Günter Schmidt
 Piotr Formanowicz
 Marta Kasprzak
 Maciej Drozdowski
 Piotr Lukasiak
 Marta Szachniuk

Rysunek 3 Przykładowa strona autora

Strona autora Google Scholar zawiera następujące informacje i dane:


- zdjęcie (jeżeli zostało uzupełnione),
- tytuł, nazwę uczelni, jednostki oraz zakresu badań,
- listę publikacji wraz z ilością cytowań oraz datą wydania,
- statystyki autora ogólne i z podziałem na lata (między innymi cytowania, h-index),
- listę współautorów prac danego autora.

Google

Scholar

←

Eksportuj



Jacek Blazewicz

Scheduling subject to resource constraints: classification and complexity

[HTML] z sciencedirect.com

Autorzy

Jacek Blazewicz, Jan Karel Lenstra, AHG Rinnooy Kan

Data publikacji

1983/1/31

Czasopismo

Discrete Applied Mathematics

Tom

5

Wydanie

1

Strony

11-24

Wydawca


North-Holland

Opis

Abstract In deterministic sequencing and scheduling problems, jobs are to be processed on machines of limited capacity. We consider an extension of this class of problems, in which the jobs require the use of additional scarce resources during their execution. A classification scheme for resource constraints is proposed and the computational complexity of the extended problem class is investigated in terms of this classification. Models involving parallel machines, unit-time jobs and the maximum completion time criterion are studied ...

Cytowania łącznie

Cytowane przez 1164



Rysunek 4 Widok publikacji

Po wejściu w wybraną publikację ukazuje się jej szczegółowy widok. Informacje w nim zawarte to między innymi nazwa, autorzy, opis, łączna liczba cytowań czy histogram cytowań w podziale na lata.

Niewątpliwą zaletą Google Scholar jest jego ogólnodostępność jak i szybkość działania. Oferuje najobszerniejszą bazę ze wszystkich istniejących rozwiązań. Niestety, przez to, że każdy może wprowadzać tam dane jakość wyników jest dosyć niska. Dlatego też Google

Scholar powinien służyć głównie celom informacyjnym. Nie powinno się klasyfikować pracowników według statystyk generowanych przez ten portal.

	ilość cytowań	h-index
prof. dr hab. inż. Jacek Błazewicz	9338	42
prof. zw. dr hab. inż. Jan Węglarz	5970	34
prof. dr hab. inż. Roman Słowiński	21891	74
dr inż. Miłosz Kadziński	584	13

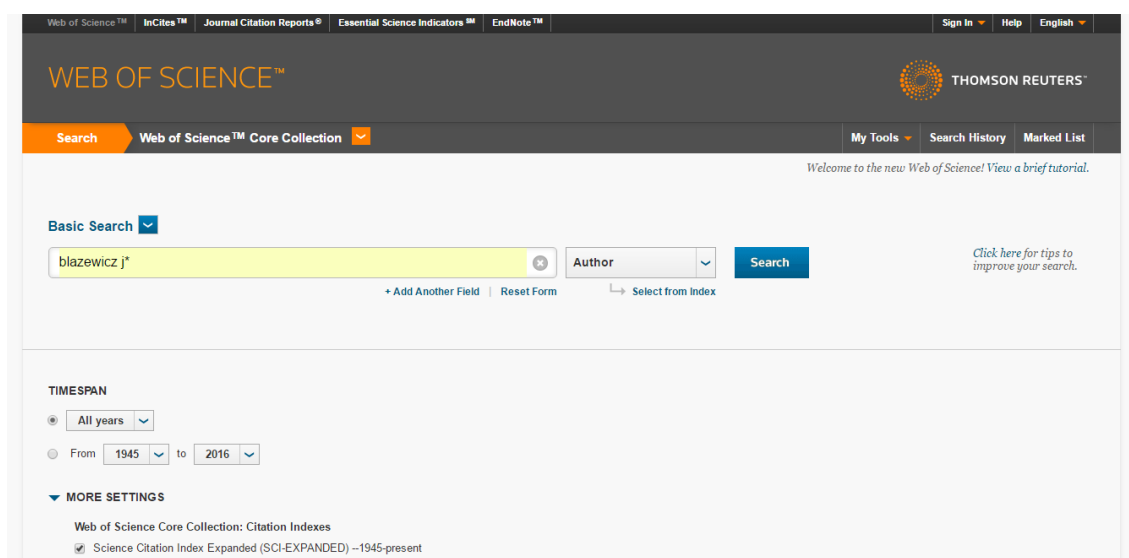
Tabela 1 Porównanie statystyk dla bazy Google Scholar

3.2 Web Of Science

Kolejnym popularnym źródłem informacji na temat publikacji naukowych jest portal Web Of Science. Obejmuje bazy danych firmy Thomson Reuters. Oferuje jeden z najobszerniejszych zbiorów publikacji naukowych: ponad 55 milionów rekordów z najważniejszych czasopism, książek i konferencji naukowych [15]. W przeciwieństwie do Google Scholar nie jest to portal darmowy. Na podstawie serwisu zwaną Core Collection składają się następujące bazy:

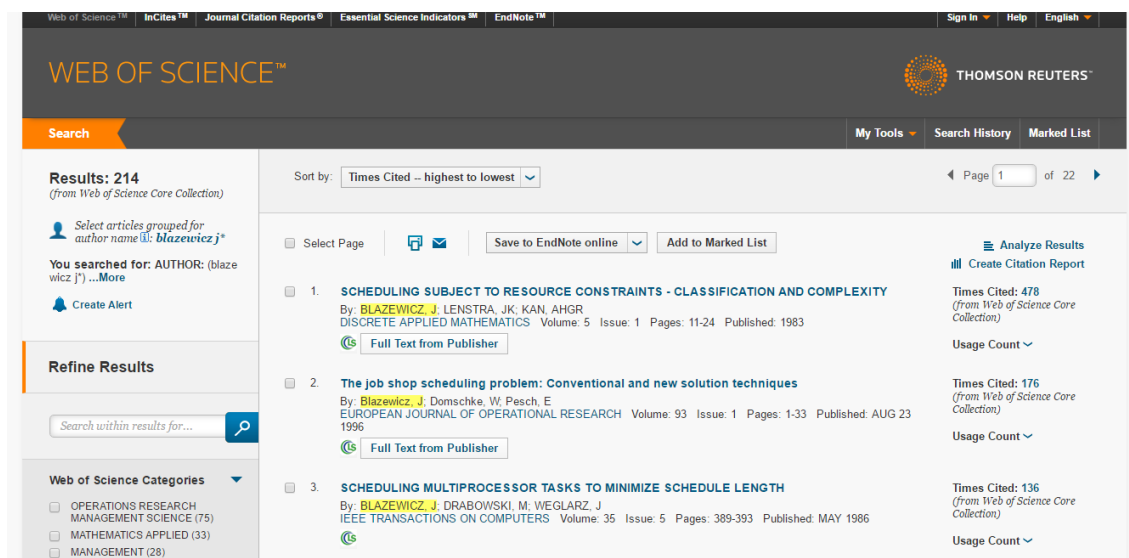
- składające się na Citation Indexes:
 - Science Citation Index Expanded – czasopisma od 1945 roku,
 - Social Sciences Citation Index – czasopisma z nauk społecznych od 1956 roku,
 - Arts & Humanities Citation Index – czasopisma z dziedziny sztuki oraz nauk humanistycznych od 1975 roku,
 - Conference Proceedings Citation Index – publikacje z konferencji od 1990 roku,
 - Book Citation Index – cytowania z książek od 2010 roku,
 - Emerging Sources Citation Index – źródła od 2015 roku.
- należące do Chemical Indexes:
 - Current Chemical Reactions – reakcje chemiczne od 2010 roku. Zawiera również dane z *National de la Propriete Industrielle structure* od 1840,
 - Index Chemicus – dane od 2010 roku.

Bazy Science Citation Index Expanded oraz Social Sciences Citation Index składają się na corocznie publikowaną listę Journal Citations Reports (JCR).



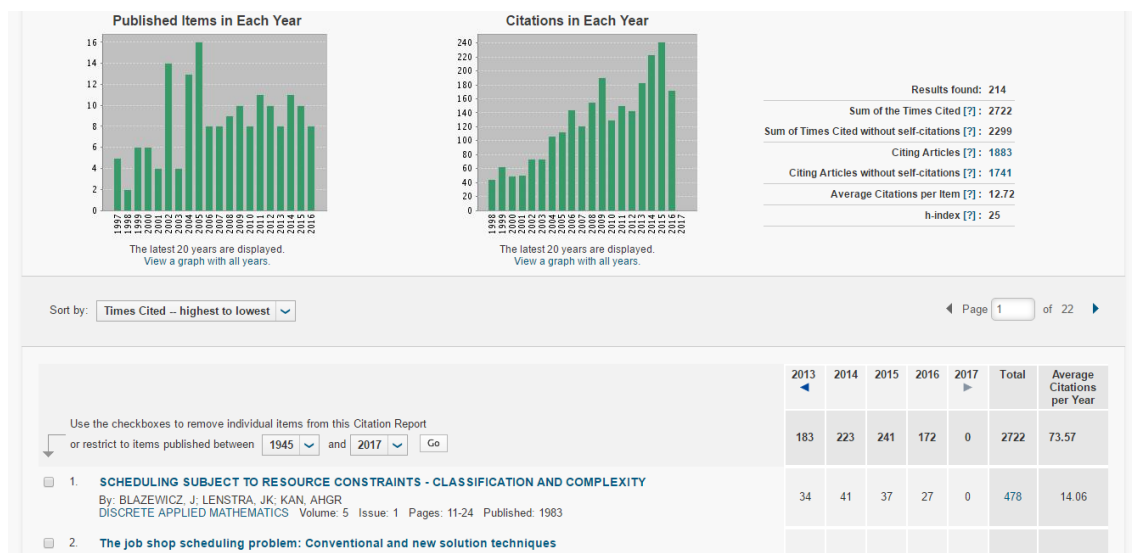
Rysunek 5 Basic Search w Web Of Science

Podstawowy typ przeszukiwania bazy danych Web Of Science nazywa się Basic Search. Dla podanego hasła zwraca listę tytułów w czasopismach, książkach i konferencjach z bazy Web Of Science wraz z cytowaniami w materiałach z tego samego zakresu. Jest to domyślny typ po wejściu na główną stronę portalu: webofknowledge.com. W odpowiednim polu należy wpisać hasło wyszukiwania oraz wybrać jego rodzaj, którym może być na przykład autor czy tytuł publikacji. Istnieje możliwość uściślenia wyszukiwania według roku wydania oraz wybrania interesujących nas baz.



Rysunek 6 Przykładowy wynik wyszukiwania

Po wyszukiwaniu wyświetla się lista znalezionych publikacji. Oprócz podstawowych informacji o danej pozycji, jak tytuł czy autorzy wyświetlają się informacje miejscu publikacji: nazwa źródła, wydanie, strony. Po prawej stronie wyświetla się liczba cytowań w bazie Web Of Science. Istnieje możliwość posortowania listy innymi po liczbie cytowań czy roku publikacji.



Rysunek 7 Widok Citation Report

Więcej informacji użytkownik może otrzymać po stworzeniu Citation Report. Po obliczeniu statystyk wyświetla się widok danego raportu. Oprócz listy publikacji zawiera on również rozdzielenie cytowań na lata (dla konkretnej pracy w tabelce oraz dla wszystkich w postaci wykresu). Oblicza również ważne z punktu widzenia oceny dorobku naukowego statystyki bibliometryczne:

- sumę cytowań bez autocytowań (cytowania swoich prac we własnych publikacjach),
- ilość cytujących artykułów,
- ilość cytujących artykułów bez autocytowań,
- średnią liczbę cytowań,
- Index Hirscha.

Za pomocą Web Of Science istnieje również możliwość wyszukiwania cytowanych pozycji bibliograficznych (Cited Reference Search). Ten sposób przeszukiwania zwraca także cytowane prace, które nie zostały opublikowane w czasopiśmie, książkach czy konferencjach zawartych w bazach Thomson Reuters.

Cited Reference Search

Find the articles that cite a person's work.

Step 2: Select cited references and click "Finish Search."

Hint: Look for cited reference variants (sometimes different pages of the same article are cited or papers are cited incorrectly).

View our Cited Reference Search tutorial.

CITED REFERENCE INDEX

References: 1 - 50 of 603

Page 1 of 13

Select Page

Select All*

Clear All

Finish Search

Select	Cited Author	Cited Work [SHOW ABBREVIATED TITLES]	Year	Volume	Issue	Page	Identifier	Citing Articles **
<input type="checkbox"/>	Adamiak, RW; Blazewicz, J; Formanowicz, P; Gdaniec, Z; Kasprzak, M; Popenda, M; Szachniuk, M [Hide all authors]	JOURNAL OF COMPUTATIONAL BIOLOGY Title: An algorithm for an automatic NOE pathways analysis of 2D NMR spectra of RNA duplexes	2004	11	1	163	DOI: 10.1089/106652704773416948	13
<input type="checkbox"/>	Antczak, M... Blazewicz, J. + [Show all authors]	Foundations of Computing and Decision Sciences Title: DomAns - pattern based method for protein domain boundaries prediction and analysis	2011	36	2			1
<input type="checkbox"/>	Antczak, Maciej... Blazewicz, Jacek + [Show all authors]	NUCLEIC ACIDS RESEARCH Title: RNApdbee-a webserver to derive secondary structures from pdb files of knotted and unknotted RNAs	2014	42	W1	W368	DOI: 10.1093/nar/gku330	9
<input type="checkbox"/>	Bacardit, J.... Blazewicz, J. + [Show all authors]	Conference: Proceedings of Genetic and Evolutionary Computation Conference (GECCO) Title: Coordination number prediction using learning classifier systems: performance and interpretability	2006			247	DOI: 10.1145/1143997.1144041	6
<input type="checkbox"/>	Bacardit, Jaume... Blazewicz, Jacek + [Show all authors]	GECCO 2006: Genetic and Evolutionary Computation Conference, Vol 1 and 2 Conference: 8th Annual Genetic and Evolutionary Computation Conference	2006			247		6

Rysunek 8 Lista wyników Cited Reference Search

Sposób wyszukiwania nie różni się praktycznie od tego z Basic Search. Różnice widać dopiero w jego wynikach, gdzie lista znalezionych rekordów jest wyświetlana w postaci tabeli. Ten typ wyszukiwania informacji o pozycjach daje więcej rezultatów, lecz obciążony jest wieloma błędami. W wielu przypadkach nie ma podanej prawidłowej nazwy cytowanej pracy. Z tego powodu jak i przez podawanie różnych stron występowania cytatu wiele pozycji zostaje rozdzielona na liście wraz z liczbą ich cytowań. Kolejnym minusem Cited Reference Search jest brak możliwości wygenerowania raportu cytowań, czyli między innymi automatycznego obliczenia statystyk (suma cytowań, indeks Hirscha).

	ilość publikacji	ilość cytowań	h-index
prof. dr hab. inż. Jacek Błazewicz	214	2722	25
prof. zw. dr hab. inż. Jan Węglarz	87	1146	19
prof. dr hab. inż. Roman Słowiński	225	5982	38
dr inż. Miłosz Kadziński	26	321	11

Tabela 2 Porównanie statystyk dla bazy WOS Basic Search

	ilość publikacji	ilość cytowań	h-index
prof. dr hab. inż. Jacek Błazewicz	613	4643	31
prof. zw. dr hab. inż. Jan Węglarz	271	2822	26
prof. dr hab. inż. Roman Słowiński	977	11606	51
dr inż. Miłosz Kadziński	46	361	11

Tabela 3 Porównanie statystyk dla bazy WOS Cited Reference Search

3.3 Scopus

Scopus to baza danych bibliograficznych prowadzona przez wydawnictwo Elsevier założona w 1995 roku. Zawiera ponad 60 milionów rekordów dotyczących czasopism oraz ponad 130 tysięcy książek [21]. Dostęp do serwisu jest płatny, lecz od 2012 jest on dostępny

bezpłatnie dla instytucji naukowych w ramach licencji krajowej udzielonej Ministerstwu Nauki i Szkolnictwa Wyższego.

Document search | **Author search** | Affiliation search | Advanced search

blażewicz x Author Initials or First Name... e.g. J.L. Q ?

Affiliation... e.g. University of Toronto

☐ Show exact matches only

ORCID... e.g. 1111-2222-3333-4444 Q

Limit to:

Subject Areas

☒ Life Sciences ☒ Physical Sciences

☒ Health Sciences ☒ Social Sciences & Humanities

Search history Combine queries... e.g. #1 AND NOT #3 Q ?

1 TITLE-ABS-KEY (blażewicz j) 1 document results

Top of page

Rysunek 9 Wyszukiwanie w bazie Scopus

Na stronie głównej www.scopus.com znajduje się wyszukiwarka. Bazę można przeszukiwać według dokumentów, autora czy jednostki naukowej lub przy użyciu wyszukiwania zaawansowanego. Tak jak na innych portalach tego typu wyszukiwanie można ograniczyć według kilku czynników, między innymi daty publikacji czy dziedziny naukowej.

Author last name "blażewicz" Edit

19 of 33 author results Show Profile Matches with One Document | About Scopus Author Identifier

Sort on: Document Count | Author (A-Z)

☐ Show exact matches only

Refine

Limit to Exclude

Source Title

☐ Proceedings Of The Biological Society Of Washington (3)

☐ Acta Of Bioengineering And Biomechanics (2)

☐ Acta Physica Polonica A (2)

☐ Annals Of Operations Research (2)

☐ BMC Bioinformatics (2)

Affiliation

☐ Uniwersytet Lodzki (3)

☐ Akademia Gorniczo-Hutnicza Im. Stanisława Staszica W Krakowie (2)

☐ Instytut Automatyki (2)

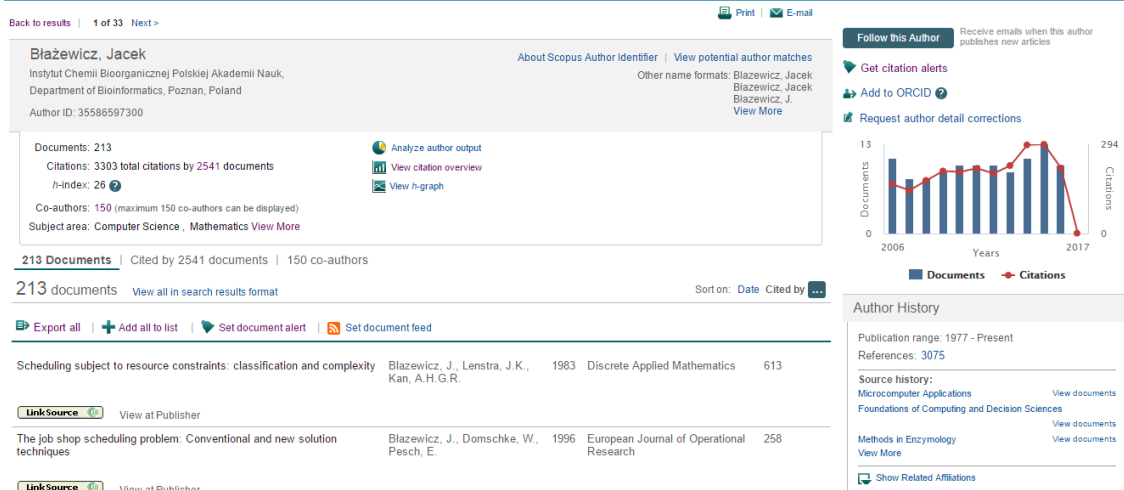
☐ Instytut Automatyki (2)

Rank	Author Name	Document Count	Subject Areas	Institution	Country
1	Blażewicz, Jacek	213	Computer Science ; Mathematics ; Decision Sciences ; ...	Instytut Chemii Bioorganicznej Polskiej Akademii Nauk	Poland
2	Blażewicz, Stanisław	77	Materials Science ; Engineering ; Chemistry ; ...	Akademia Gorniczo-Hutnicza Im. Stanisława Staszica W Krakowie	Poland
3	Blażewicz, Marta	57	Chemistry ; Materials Science ; Engineering ; ...	Akademia Gorniczo-Hutnicza Im. Stanisława Staszica W Krakowie	Poland
4	Blażewicz-Paszkowycz, Magdalena	50	Agricultural and Biological Sciences ; Earth and Planetary Sciences ; Environmental Science ; ...	Uniwersytet Lodzki	Poland

Rysunek 10 Wyszukani autorzy

Po wyszukaniu interesującego nas autora wyświetla się lista twórców pasujących do wpisanego zapytania. Na liście oprócz imienia i nazwiska widoczny jest zakres badań oraz jednostka organizacyjna. W celu ułatwienia przeszukiwania zestawiania możliwe jest jego przefiltrowanie.

Author details

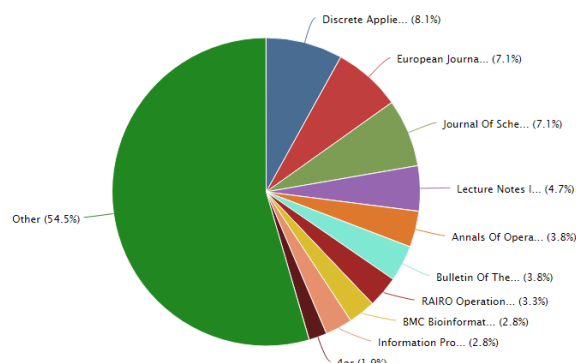


Rysunek 11 Szczegóły autora w bazie Scopus

W widoku szczegółowych informacji o autorze widoczna jest lista jego publikacji dostępna w bazie Scopus. Istnieje możliwość podejrzenia listy artykułów cytujących danego autora oraz współautorów. Serwis automatycznie oblicza h-index oraz rysuje wykres liczby prac i cytowań w zależności od roku. Po wejściu w konkretną publikację wyświetlona zostają jej szczegółowe informacje, na przykład streszczenie, lista cytowań.

Source	Documents
Discrete Applied Mathematics	17
European Journal Of Operational ...	15
Journal Of Scheduling	15
Lecture Notes In Computer Scienc...	10
Annals Of Operations Research	8
Bulletin Of The Polish Academy Of...	8
RAIRO Operations Research	7
BMC Bioinformatics	6
Information Processing Letters	6
4or	4
Computational Biology And Chem...	4
International Journal Of Applied M...	4
Nucleic Acids Research	4
Acta Biochimica Polonica	3
Archivum Automatyki I Telemecha...	3
Bioinformatics	3

Documents by source



Rysunek 12 Wykres dokumentów

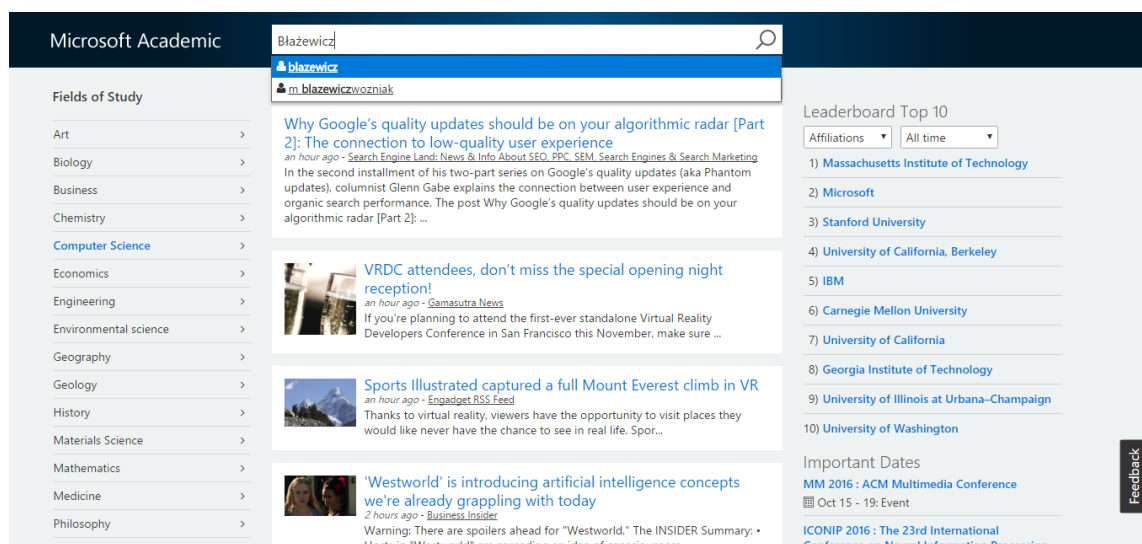
Portal Scopus udostępnia jeszcze inną ciekawą funkcję: generowanie i wyświetlanie grafów statystyk. Na rysunku 12 widać wykres dokumentów według miejsca publikacji. Istnieje możliwość wygenerowanie podobnego grafu według typów publikacji, roku oraz przedmiotu badań. Użytkownik może wyświetlić również wykres indeksu Hirscha, cytowań według lat oraz współautorów.

	ilość publikacji	ilość cytowań	h-index
prof. dr hab. inż. Jacek Błazewicz	213	3305	26
prof. zw. dr hab. inż. Jan Węglarz	89	1118	20
prof. dr hab. inż. Roman Słowiński	264	8061	44
dr inż. Miłosz Kadziński	31	366	11

Tabela 4 Porównanie statystyk dla bazy Scopus

3.4 Microsoft Academic

Microsoft Academic to wyszukiwarka danych bibliograficznych prowadzona przez amerykańską firmę Microsoft. Jest siecią semantyczną zawierającą metadane prac naukowych publikowanych w czasopismach i na konferencjach. W lutym 2014 serwis indeksował ponad 39,9 milionów publikacji oraz 19.9 milion autorów [9]. Powstał po zamknięciu w 2012 starego portalu o nazwie Microsoft Academic Search i włączeniu go do projektu wyszukiwarki internetowej Bing. Zmiana ta została wymuszona przez potrzebę dostosowania serwisu do nowoczesnych przeglądarek internetowych [8].



Rysunek 13 Wyszukiwanie w Microsoft Academic

Na pierwszej stronie wyszukiwarki zaprezentowana została lista popularnych artykułów. Istnieje możliwość wyszukania ich po konkretnych kategoriach. Dodatkowo wyświetlana jest lista najpopularniejszych instytucji oraz aktualnych konferencji. Microsoft Academic umożliwia zatem przeszukiwania zaindeksowanych publikacji nie tylko po nazwisku autora, ale również po jednostce organizacyjnej, dziedzinie naukowej czy konferencji, podczas której praca została zaprezentowana.

Rysunek 14 Lista wyszukanych pozycji w Microsoft Academic

Po wyszukaniu konkretnego autora czy tematu publikacji wyniki zostają zaprezentowane w postaci czytelnej listy. Można dane publikacje przefiltrować między innymi po jednostce naukowej, dziedzinie nauki czy nazwisku autora. Dodatkowo przedstawiona jest lista zaindeksowanych w portalu autorów, którzy pasują do wyszukiwanego hasła. Niestety, jedyną statystyką wyświetlaną po wyszukiwaniu jest łączna liczba cytowań prac. W celu wyznaczenia bardziej zaawansowanych metryk, jak chociażby indeksu Hirscha należy wykorzystać inne narzędzie.

	ilość publikacji
prof. dr hab. inż. Jacek Błażewicz	340
prof. zw. dr hab. inż. Jan Węglarz	137
prof. dr hab. inż. Roman Słowiński	390
dr inż. Miłosz Kadziński	34

Tabela 5 Porównanie ilości publikacji poszczególnych autorów w Microsoft Academic

3.5 Scholar+

Na koniec warto wspomnieć o projekcie Scholar+. Jest to portal zaimplementowany w ramach pracy inżynierskiej w 2014 roku przez studentów Politechniki Poznańskiej Tomasza Charońskiego, Monikę Grabowską, Mateusza Lango oraz Jakuba Lewińskiego, promotorem pracy był dr hab. inż. Jerzy Nawrocki. Powstał on w celu wyeliminowania niedogodności istniejących narzędzi wykorzystywanych do przeglądania danych bibliograficznych i generowania na ich podstawie statystyk. W założeniu wprowadzał również nowe, nieistniejące w tamtym czasie funkcjonalności. Umożliwiał filtrowanie danych w systemie, pozwalając na oddzielenie publikacji i cytowań pochodzących z innych źródeł niż artykuły czy książki znajdujące się na listach punktowanych. Rozwiązywał też problem obliczania wskaźników dla grup naukowców oraz prac autorów publikujących dla wielu jednostek. Dzięki

zaimplementowaniu mechanizmu synonimów system łączył magazyny znajdujące się w Google Scholar, które zostały różnie opisane [13].

System bazował na danych zaciągniętych z serwisu Google Scholar. Aby dany autor albo jednostka organizacyjna znalazła się w serwisie, administrator musiał wprowadzić w panelu administracyjnym link z pierwotnej strony. System sam pobierał dane i zapisywał w lokalnej bazie danych. Za pomocą interfejsu użytkownika możliwe było zarządzanie organizacjami (między innymi przypisywanie użytkowników lub zmiana podstawowych informacji) czy import list czasopism punktowanych [12].

Po wyszukaniu konkretnego autora wyświetlały się informacje o nim. Z podstawowych informacji zaprezentowane było zdjęcie, imię i nazwisko, jednostka naukowa oraz obszary zainteresowań naukowca. System wyświetla listę publikacji, którą użytkownik mógł przefiltrować po źródle i latach wydania czy cytowań. Aplikacja od razu wyliczała statystyki autora, takie jak liczba cytowań, h-index, liczba publikacji, Impact Factor oraz punkty ministerialne.

Widok wyświetlał listę pracowników, którą można było posortować po indeksie Hirscha czy cytowaniach. Ponad zestawieniem przedstawione były informacje o jednostce:

- nazwa,
- logo,
- adres,
- liczba organizacji,
- liczba pracowników,
- cytowania,
- h-index,
- publikacje,
- impact factor,
- punkty ministerialne.



Jerzy Błaszczyński

Poznań University of Technology

Obszar(y) badań: artificial intelligence, decision support, machine learning, preference learning, multi-criteria decision aid

[Profil Google Scholar](#)

Indeks cytowań

Cytowania	381
h-index	9
Publikacje	39
Impact Factor	13.345
Punkty Ministerstwa	109

Filtrowanie

Źródło publikacji

- ☒ Google Scholar
- ☐ JCR+CPCI
- ☐ tylko JCR

Pomiędzy latami

 -

Źródło cytowań

- ☒ Google Scholar
- ☐ JCR+CPCI
- ☐ tylko JCR

Pomiędzy latami

 -

Filtruj

Tytuł i autorzy	▼ Liczba cytowań	Impact Factor	Punkty Ministerstwa	Rok
Multi-criteria classification—A new scheme for application of dominance-based decision rules Jerzy Błaszczyński, Salvatore Greco, Roman Słowiński	112	1.096	9	2007
Sequential covering rule induction algorithm for variable consistency rough set approaches CPCI Jerzy Błaszczyński, Roman Słowiński, Marcin Szeląg	49	2.833	20	2011
Monotonic variable consistency rough set approaches Jerzy Błaszczyński, Salvatore Greco, Roman Słowiński, Marcin Szeląg	43	2.09		2009

Rysunek 15 Widok autora

Informacje

Liczba organizacji	3
Liczba pracowników	27
Cytowania	33307
h-index	86
Publikacje	1751
Impact Factor	291.218
Punkty Ministerstwa	1560

Filtrowanie

Autorzy / Organizacje	▼ Cytowania	h-index
Roman Słowiński	15868	64
Jacek Błazewicz	6649	36

Rysunek 16 Widok organizacji

3.6 Podsumowanie

Przejrzenie istniejących baz bibliograficznych uwydatniło różnice między nimi. Poszczególne narzędzia wyróżniają się interfejsem graficznym, dostępną funkcjonalnością, wyświetlanymi statystykami oraz przede wszystkim wielkością i jakością bazy.

	ilość publikacji	ilość cytowań	h-index
Scopus	213	3305	26
Google Scholar	474	9338	42
WOS basic search	214	2722	25
WOS cited reference search	613	4643	31
Microsoft Academic	340	brak danych	brak danych

Tabela 6 Porównanie danych pochodzących z różnych serwisów dla
prof. dr hab. inż. Jacka Błazewicza

	Liczba publikacji	Suma cytowań	H-index
Scopus	264	8061	44
Google Scholar	501	21891	74
WOS basic search	225	5982	38
WOS cited reference search	977	11606	51
Microsoft Academic	390	brak danych	brak danych

Tabela 7 Porównanie danych pochodzących z różnych serwisów dla
prof. dr hab. inż. Romana Słowińskiego

	Liczba publikacji	Suma cytowań	H-index
Scopus	31	366	11
Google Scholar	38	584	13
WOS basic search	26	321	11
WOS cited reference search	46	361	11
Microsoft Academic	34	brak danych	brak danych

Tabela 8 Porównanie danych pochodzących z różnych serwisów dla
dr inż. Miłosza Kadzińskiego

4. Projekt wzbogacenia funkcjonalności baz bibliometrycznych

Analiza istniejących narzędzi bibliometrycznych pod kątem wielkości bazy oraz oferowanych funkcjonalności podsunęła pomysł stworzenia serwisu, który niwelowałby ich niedogodności bazując na informacjach przechowywanych przez te portale. Jednym z podstawowych problemów pojawiających się w istniejących rozwiązaniach było rozdzielanie publikacji. Cytowania tej samej pracy często były przechowywane w bazie danych jako wiele rekordów spowodowane różnicami w ich opisach w cytujących artykułach. Kolejnym problemem, który należało rozwiązać były niekompletne bazy czasopism naukowych, konferencji i książek. Oczywiście większość źródeł się pokrywa, lecz nadal występuje wiele prac wydawanych w wartościowych publikacjach, które możemy znaleźć tylko w niektórych z dostępnych baz.

4.1 System Scholar+

Po zapoznaniu się z pracą inżynierską dotyczącą systemu Scholar+ uznałem, że najlepszym sposobem realizacji założeń będzie przejęcie tego projektu. Portal rozwiązywał większość niedogodności związanych z istniejącymi narzędziami oraz implementował wiele funkcjonalności, które rozszerzały pierwotne założenia oraz zwiększały użyteczność i przydatność serwisu.

Realizacja pracy miała zatem odbyć się w następujących krokach:

1. Uruchomienie systemu na serwerze: Projekt inżynierski Scholar+ został oddany w 2014 roku i od tego czasu nie był rozwijany ani aktualizowany. Wiele technologii z których korzystał zostało zaktualizowane. Aby aplikacja była sprawna, przynajmniej w podstawowym stopniu, należało korzystać z większości starszych narzędzi. Odpowiedni serwer miała zapewnić uczelnia.
2. Zaktualizowanie kodu pod względem zmienionego otoczenia, tak aby w pełni realizował pierwotne funkcje: Z powodu braku udostępnionego API do interakcji z Google Scholar działanie systemu opierało się na pobieraniu z portalu całych stron. Następnie z kodu HTML za pomocą techniki *web scrappingu* (technika komputerowa wydobywania informacji ze stron internetowych [17]) wyciągane były potrzebne rekordy. Tak przygotowane dane system wprowadzał do bazy, z którą komunikował się interfejs użytkownika. Po wstępnej analizie działania systemu na pierwszy plan wysuwały się dwa problemy do rozwiązania:
 - o kwestia blokowania przez system Google'a automatu pobierającego dane,

- zmieniony interfejs użytkownika, a co za tym idzie różnice w znacznikach elementów w kodzie HTML, potrzebnych do odpowiedniego parsowania strony.

Pierwszy z tych problemów autorzy aplikacji rozwiązyali stosując klienta sieci TOR (sieć zapobiegająca analizie ruchu sieciowego, przez co zapewnia użytkownikowi prawie anonimowy dostęp do sieci [14]). Za każdym razem, gdy podczas pobierania danych Google blokował użytkownika i wymuszał wpisanie kodu CAPTCHA, automat resetował klienta TOR. Przy następnej próbie ściągnięcia ostatniej strony serwer zwracał wynik bez błędów. Kwestie zmieniającego się wyglądu strony programiści rozwiązyali poprzez zapisywanie znaczników na serwerze. Aplikacja łączyła się z lokalnym serwerem bazy danych, gdzie w odpowiedniej tabeli przechowywane były ciągi znaków za pomocą których biblioteka obsługująca web scrapping (Selenium) mogła wydobyć potrzebne informacje.

W tym punkcie realizacji pracy należało przetestować pobieranie stron przez automat oraz zaktualizować odpowiednie dane w tabeli konfiguracyjnej.

3. Zaktualizowanie bibliotek i technologii: Aby dało się odpowiednio i bez problemu rozwijać system niezbędne było zaktualizowanie narzędzi do najnowszych wersji. Wiele bibliotek, jak na przykład służące do obsługi bazy danych Hibernate i PostgreSQL, zostało wiele razy zaktualizowanych od zakończenia prac nad Scholar+. Aplikacja była napisana w języku Java przy użyciu framework'a Ninja, który został wybrany przede wszystkim ze względu na swoją lekkość. Technologia jest nadal rozwijana (ostatnia wersja została opublikowana 27 sierpnia 2016 roku), ale istnieje obecnie wiele innych rozwiązań, które spełniałyby wymagania. Niezbędnym przy takiej aplikacji internetowej wydaje się całkowite rozdzielenie warstwy interfejsu użytkownika od logiki biznesowej, czego obecnie używany Framework niestety nie zapewniał.
4. Dodanie nowej funkcjonalności: System Scholar+ jest dobrą podstawą do stworzenia bardziej kompleksowego i dostosowanego do wymagań użytkownika portalu. Za tę część realizacji rozwiązania (wymagania funkcjonalne, projekt, implementacja) odpowiedzialna miała być grupa projektowa w ramach przedmiotu Software Development Studio w składzie inż. Dominik Bargowski oraz inż. Maciej Kasprzycki. Opiekunem zespołu jest mgr inż. Sylwia Kopczyńska.

Uruchomienie systemu w surowej postaci (bez danych) na serwerze uczelnianym nie okazało się bardzo trudnym zadaniem, ale oczywiście niepozbawionym pewnych komplikacji. Najważniejszym problemem, na który trzeba było być wyczulonym to instalacja odpowiednich

bibliotek i narzędzi. Na szczęście udało się skorzystać z najnowszej wersji systemu Ubuntu 14.04.4 oraz środowiska uruchomieniowego Tomcat 7.0.52. Gdy już strona internetowa była dostępna spoza sieci Politechniki Poznańskiej można było przejść do aktualizacji systemu, aby prezentował aktualne dane.

Zaktualizowanie w lokalnej bazie danych znaczników służących do wyciągania danych z kodu HTML nie przysporzyło większych problemów. W celu sprawdzenia ścieżki (tak zwanego XPath'a) danego elementu na stronie można nawet skorzystać z narzędzi deweloperskich przeglądarki Google Chrome. Trudności nastąpiły w momencie testów pobierania danych z Google Scholar. Podczas etapu aktualizowania informacji przez automat, w którym wykorzystywany był klient TOR (co było wymagane, ponieważ w krótkim czasie system wysyłał wiele żądań do serwera) okazało się, że Google zaimplementował nowe filtry. Obecnie potrafi wykryć oraz blokować klienta korzystającego z sieci TOR. Po wielu próbach rozwiązania tego problemu wraz z grupą projektową SDS doszliśmy do wniosku, że należy niestety zaniechać dalszych prób ożywienia i rozwinięcia systemu Scholar+ i należy skupić się na stworzeniu własnego, unikalnego projektu.

4.2 Analiza możliwości wzbogacenia istniejących serwisów

Po nieudanym przywróceniu portalu Scholar+ należało zaprojektować rozwiązanie od nowa. Istotne w tym momencie pracy było przeanalizowanie istniejących baz i portali bibliometrycznych pod względem możliwości ich rozszerzania.

Oparcie nowego systemu o Google Scholar byłoby najlepszym rozwiązaniem. Serwis udostępnia wiele informacji, ma bardzo rozbudowaną bazę publikacji i cytowań. Zaimplementowanie w nowym serwisie mechanizmów łączenia rozdzielonych publikacji rozwiązałoby problem złych opisów cytowań. Dodanie spisu czasopism punktowanych i filtra publikacji według nich zniwelowałoby niedogodności związane z ogólnodostępnością portalu (a co za tym idzie niskiej jakości cytowań). Niestety, jak to zostało pokazane w poprzednim rozdziale, Google Scholar zawiera jedną dużą wadę. Nie udostępnia żadnego systemu w postaci API, za pomocą którego byłoby można pobierać dane na potrzeby własnego rozwiązania. Jak pokazał wcześniejszy rozdział, również wyciąganie danych za pomocą web scrappingu nie daje pożądanych rezultatów, ponieważ Google blokuje automatyczne pobieranie stron.

Po odrzuceniu Google Scholar'a jako bazy dla nowego portalu wybór padł na portal Web Of Science, prowadzony przez firmę Thomson Reuters. Przed rozpoczęciem prac nad właściwym systemem najpierw należało dokładnie przeanalizować możliwości i ograniczenia danej bazy danych.

Web Of Science udostępnia API [16], istnieje zatem możliwość bezproblemowego pobierania danych z bazy. Aby korzystać z web service'u wymagane jest posiadanie wykupionego dostępu do serwisu, a zalogować się można na dwa sposoby:

- za pomocą loginu i hasła, wysłanego przez pracownika pomocy technicznej Web Of Knowledge,
- wysyłając zapytania z adresu IP, który ma nieograniczony dostęp do portalu (na przykład z sieci uczelnianej).

Przy pomocy API można wysłać dokładnie takie same zapytanie, co przy pomocy strony internetowej. Istnieje zatem możliwość wyszukania publikacji między innymi według autora, tytułu, jednostki organizacyjnej, czasopisma czy daty publikacji. Metoda „serach” zwraca szczegółowe rekordy pasujące do wybranej wartości. Zatem informacje jakie użytkownik może otrzymać to między innymi: autorzy czy nazwa, miejsce i liczba cytowań publikacji. Za pomocą innych udostępnionych metod można pobrać:

- prace cytujące daną publikację (metoda „citingArticles”),
- prace cytowane przez daną publikację (metoda „citedReferences”),
- powiązane prace, czyli posiadające te same cytowania (metoda „relatedRecords”).

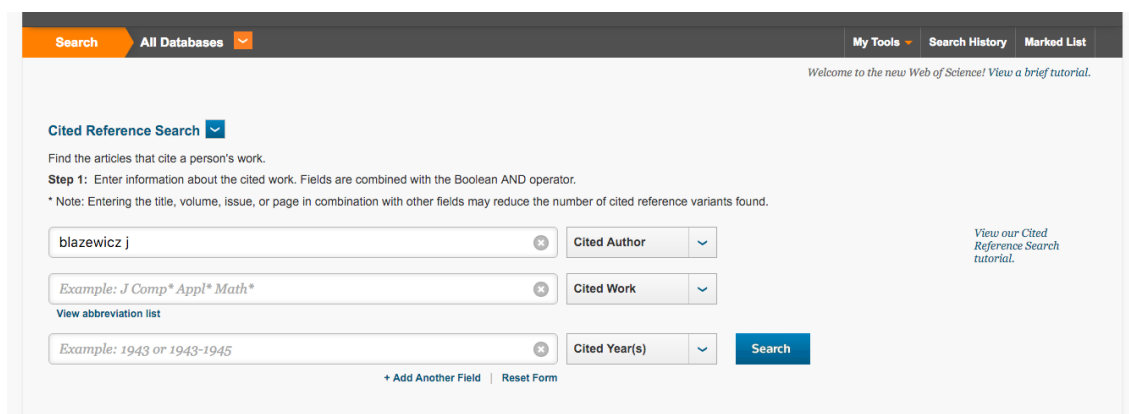
API Web Of Science umożliwia użytkownikowi przeszukiwanie wszystkich edycji bazy danych dostępnych w serwisie (między innymi Arts & Humanities Citation Index, Science Citation Index Expanded, Index Chemicus). Limit nałożony na wyszukiwanie danych to 2 zapytania na sekundę oraz maksymalnie 100 rekordów zwracanych przez pojedynczą odpowiedź systemu. Niestety, jedynym dostępnym typem wyszukiwania dostępnym w interfejsie programistycznym jest Basic Search, który umożliwia znalezienie prac publikowanych w czasopismach z listy Thomson Reuters. Przy tej opcji portal jest bardzo dobrze zaprojektowany, nie występują tu problemy znane między innymi w Google Scholar. Nie ma źle opisanych cytowań, wszystkie pisma czy konferencje w bazie WOS należą do listy ministerialnej. Nie ma za bardzo możliwości rozszerzenia tego narzędzia. Dużo bardziej interesującym z punktu widzenia nowego portalu byłoby oparcie się na Cited Reference Search.

Typ wyszukiwania Cited Reference Search zwraca cytowane prace publikowane nie tylko w czasopismach, konferencjach czy książkach z listy Thomson Reuters, ale wszystkie prace, które zostały w powyższych źródłach zacytowane. Zwraca zatem dużo więcej wyników dla pojedynczego wyszukiwania, co doskonale widać w tabeli 2 oraz w tabeli 3. Liczba publikacji zwracanych dla tego typu przeszukiwania jest nawet trzykrotnie wyższa niż w przypadku Basic Search. Dodatkowo niedoskonałości tej funkcji Web Of Science zostawiają duże pole do poprawek i usprawnień. Cytowania bywają niepoprawnie opisane oraz pochodzą często z mało wartościowych źródeł. Można tutaj zaimplementować mechanizm łączenia publikacji oraz filtr czasopism punktowanych. Cited Reference Search nie generuje nawet podstawowych statystyk (suma cytowań, średnia cytowań, indeks Hirscha). Z tych powodów oparcie własnego rozwiązania o tę opcję przeszukiwania bazy Web Of Science miałyby bardzo dużą wartość biznesową. Niestety, API portalu nie udostępnia wyszukiwania cytowanych

pozycji bibliograficznych. Jedyną możliwością, aby pobierać i wyświetlać te dane jest web scrapping. Na szczęście, próba wysłania wielu żądań w krótkim czasie na stronę główną Web Of Science nie zwróciła błędów ani niespodziewanych wyników. Dlatego też, projekt poszedł w kierunku czerpania z wyników wyszukiwania Cited Reference Search w Web Of Science.

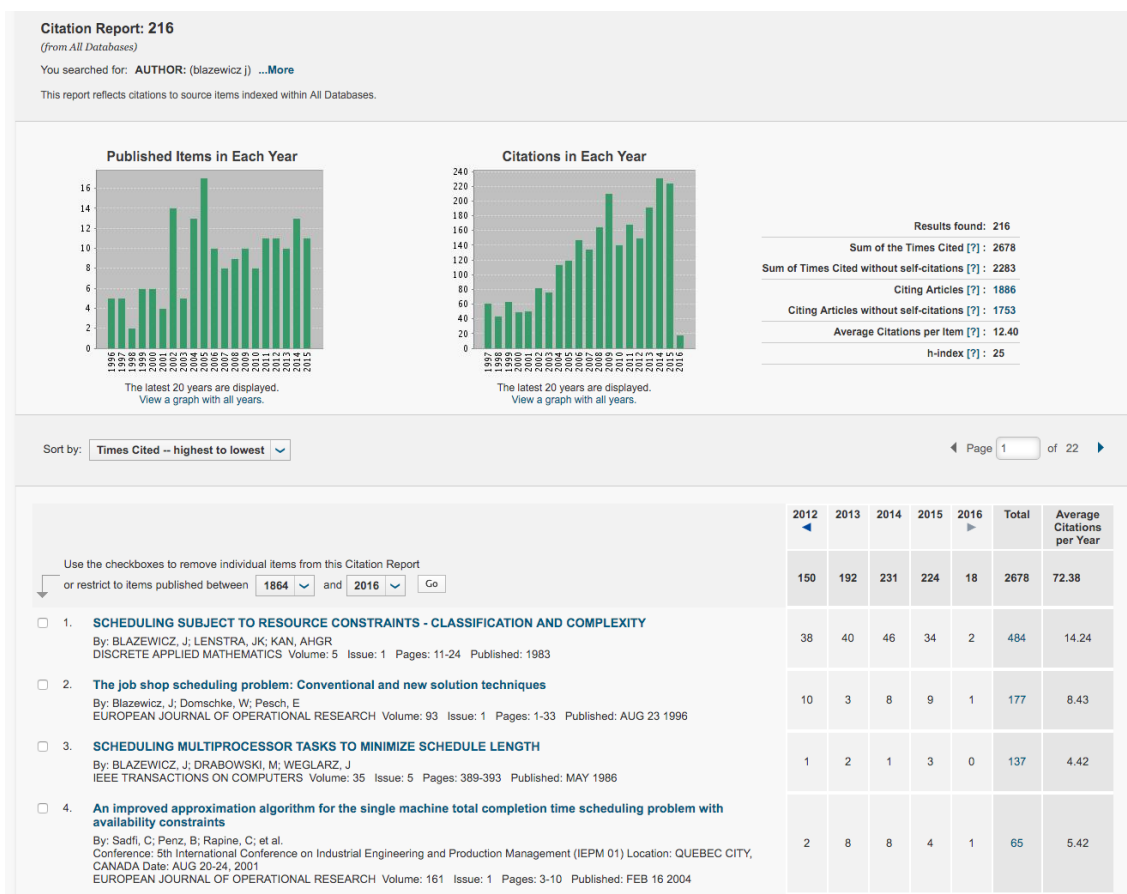
4.3 Opis założeń Web Of Science +

Projekt Web Of Science+ powinien zostać zrealizowany jako aplikacja internetowa napisana w języku angielskim. W początkowej fazie rozwoju nie musi być dostępna spoza sieci uczelnianej. Strona internetowa powinna działać w oparciu o dane z wszystkich dostępnych baz danych w serwisie Web Of Science (opcja „All Databases”) w wyniku na zapytanie Cited Reference Search.

The screenshot shows the 'Cited Reference Search' interface on the Web of Science website. At the top, there is a navigation bar with 'Search' and 'All Databases' (with a dropdown arrow). To the right are links for 'My Tools', 'Search History', and 'Marked List'. Below the navigation bar, a welcome message reads: 'Welcome to the new Web of Science! View a brief tutorial.' The main section is titled 'Cited Reference Search' with a dropdown arrow. It includes instructions: 'Find the articles that cite a person's work.' and 'Step 1: Enter information about the cited work. Fields are combined with the Boolean AND operator.' A note states: '* Note: Entering the title, volume, issue, or page in combination with other fields may reduce the number of cited reference variants found.' There are three input fields on the left, each with a plus icon: the first contains 'blazewicz j', the second has an example 'Example: J Comp* Appl* Math*' and a 'View abbreviation list' link, and the third has an example 'Example: 1943 or 1943-1945'. To the right of these are three dropdown menus: 'Cited Author', 'Cited Work', and 'Cited Year(s)'. A blue 'Search' button is located to the right of the 'Cited Year(s)' dropdown. At the bottom, there are links for '+ Add Another Field' and 'Reset Form'. On the far right, a link says 'View our Cited Reference Search tutorial.'

Rysunek 17 Przeszukiwanie All Databases w trybie Cited Reference Search

Aplikacja powinna wymagać od użytkownika wpisania jedynie nazwiska i pierwsze litery imienia szukanego autora.



Rysunek 18 Citation Report

Widok wyników wyszukiwania powinien zawierać te same informacje co Citation Report. Musi się jednak różnić od oryginalnej strony grafiką, aby nie zostać posądzonym o nielegalne kopiowanie. WOS+ będzie liczyć i wyświetlać następujące statystyki:

- Items found – liczba znalezionych rekordów,
- Sum of the Times Cited – łączna suma cytowań,
- Average Citations per Item – średnia liczba cytowań,
- h-index – indeks Hirscha.

Z powodu braku dostępnych danych nie będą wyświetlane statystyki z dopiskiem „without self-citations” (czyli bez autocytaowań) oraz statystyki artykułów cytujących. Nie będzie również opcji wyświetlania danych z rozbiem na lata oraz rysowania na ich podstawie wykresów.

Na liście wyników wyszukiwania aplikacja powinna wyświetlać pełne nazwy tytułów (czyli po kliknięciu „SHOW EXPANDED TITLES” w tabeli z WOS). Użytkownik powinien mieć możliwość posortowania rekordów po dacie publikacji oraz ilości cytowań. Na jednej stronie wyświetlane powinno być 50 wyników. W tabeli po ostatniej publikacji wliczającej się do indeksu Hirscha powinna być wyświetlana pogrubiona linia.

W początkowej fazie projektu dane mają być na bieżąco pobierane z oryginalnej strony Web Of Science. W przyszłości należałoby dodać funkcjonalność zapisywania wyników do

lokalnej bazy danych. Spowodowałoby to zdecydowane szybsze działanie całego serwisu. Byłoby również początkiem do implementacji mechanizmu scalania znalezionych duplikatów.

5. Etapy projektu

Badanie problemu zbierania danych bibliograficznych uwydatniło istotę bibliometrii w ocenie dorobku naukowego. Przegląd dostępnych narzędzi katalogujących publikacje i cytowania ukazały możliwości ich rozszerzenia. To wszystko zainspirowało do stworzenia własnego rozwiązania, którego przydatność dla pracowników naukowych wydziału byłaby znacząca.

5.1 Wymagania funkcjonalne

Po analizie możliwości serwisu Web Of Science i potrzeb użytkowników została zdefiniowana następująca lista wymagań funkcjonalnych systemu:

- działanie w oparciu o dane zwracane przez bazę All Databases w wyniku na zapytanie Cited Reference Search na portalu Web Of Science,
- prezentowanie zwracanych danych w formie tabeli,
- operowanie na pełnych nazwach czasopism,
- wyświetlanie statystyk cytowań: suma cytowań, średnia liczba cytowań oraz indeks Hirscha,
- strona napisana w języku angielskim,
- wymaganie od użytkownika podania jedynie nazwiska i pierwszej litery imienia szukanego autora,
- umożliwienie sortowania zarówno po roku publikacji i po liczbie cytowań (domyślnie po liczbie cytowań malejąco),
- grupowanie wyników po 50 pozycji na stronie,
- wyświetlanie pogrubionej linii po ostatniej publikacji wliczającej się do indeksu Hirscha.

5.2 Wykorzystane technologie

Głównym założeniem projektu było zaimplementowanie go w formie strony internetowej. Wymagało użycia którejś z dostępnych technologii umożliwiającej tworzenie rozbudowanych aplikacji internetowych. Po analizie możliwych rozwiązań wybór padł na technologię Microsoft'u: ASP.NET MVC w wersji 5. Jest to platforma aplikacyjna oparta na wzorcu Model-View-Controller, który dzieli aplikacje na trzy główne części:

- model: reprezentuje pewien problem lub logikę biznesową,
- widok: opisuje jak wyświetlić pewną część modelu w interfejsie użytkownika,
- kontroler: przyjmuje dane wejściowe od użytkownika i reaguje na jego poczynania, zarządzając aktualizacje modelu oraz odświeżenie widoków [10].

Tworzenie aplikacji w technologii ASP.NET MVC 5 opiera się zatem na następujących elementach:

- utworzenie klas w języku C# odpowiadających za dostęp do danych i operacje na nich. Do ich obsługi można wykorzystać technologię Entity Framework (technologia mapowania obiektowo-relacyjnego ORM). Jest to warstwa modelu.
- w warstwie widoku utworzenie plików o rozszerzeniu cshtml. Zawierają one kod HTML'owy wraz z odpowiednimi znacznikami zaczynającymi się od znaku „@”, które umożliwiają wyświetlanie pewnych części modelu oraz wykorzystywanie składni języka C#.
- napisanie do każdego widoku klasy kontrolera. Będzie ona odpowiadać za obsługę zdarzeń zgłoszonych przez użytkownika oraz pobierać i operować na modelu wyświetlanym przez widok.

Technologia ASP.NET MVC cechuje się między innymi:

- oparciem modelu programistycznego na interfejsach,
- elastycznym mechanizmem mapowania adresów URL [2].

W projekcie został wykorzystany .NET Framework w wersji 4.5.2. Metody tej platformy były wykorzystywane między innymi do wysyłania żądań na stronę Web Of Science. Do web scrappingu pobranych stron HTML wykorzystana została biblioteka HtmlAgilityPack, natomiast Framework'iem CSS do zaimplementowania interfejsu graficznego użytkownika był Bootstrap, rozwijany przez programistów Twitter'a. Projekt został zrealizowany za pomocą środowiska programistycznego firmy Microsoft: Visual Studio 2015.

5.3 Projekt

Tak jak to zostało wspomniane wcześniej projekt ma formę strony internetowej napisanej w języku angielskim. Aplikacja zawiera dwa widoki:

- stronę startową,
- listę wyników wyszukiwania.

Na środku strony startowej jest wyświetlane pole wyszukiwarki. Po wpisaniu przez użytkownika interesującego go hasła (posiadającego taką samą formę jak wyszukiwanie na oryginalnej stronie Web Of Science) i kliknięciu przycisku „Search” następuje przekierowanie do widoku wyników wyszukiwania. Na początku wyświetla się informacja o pobieraniu danych wraz z odpowiednią animacją. Po ukończonym wyszukiwaniu przedstawiona zostaje lista znalezionych rekordów. W poszczególnych wierszach wyświetlane są następujące informacje:

1. Liczba porządkowa.
2. Informacje na temat publikacji dostępne w wynikach wyszukiwania na stronie Web Of Science: nazwa, autorzy, miejsce publikacji, rok, wydanie, strony cytatu (jeżeli któraś z informacji nie będzie dostępna zostanie wyświetlony komunikat „N/D” oznaczający brak danych – no data).
3. Rok wydania.

4. Liczba cytowań.

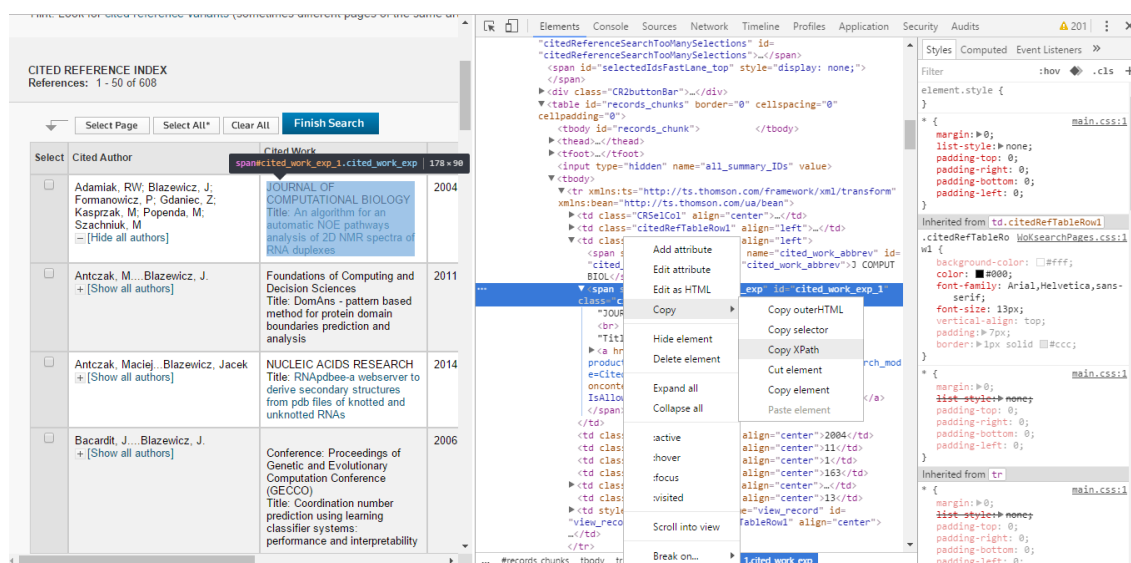
Użytkownik może posortować wyniki po wartościach z ostatnich dwóch kolumn. W tabeli po ostatniej pozycji wliczającej się do indeksu Hirscha wyświetlona zostaje gruba czarna linia.

Ponad wynikami wyszukiwania przedstawione są następujące informacje:

- wyszukiwane hasło,
- liczba publikacji,
- suma cytowań znalezionych publikacji,
- średnia liczba cytowań na publikację,
- indeks Hirscha.

Wyniki wyszukiwania są rozdzielone na kilka stron, po 50 wierszy każda. Numery stron wraz z linkami do nich wyświetlają się na dole strony.

Projekt nie zakłada buforowania danych w lokalnej bazie, dlatego informacje pobierane ze strony Web Of Science są na bieżąco ściągane, parsowane i wyświetlane użytkownikowi. Do wyciągania istotnych informacji z kodu HTML konkretnej strony służy biblioteka HtmlAgilityPack, przy pomocy odpowiednio zdefiniowanych znaczników, tak zwanych XPath. Do ich określenia wystarczy użyć narzędzia deweloperskiego dostępnego w przeglądarce internetowej Google Chrome.



Rysunek 19 Wyciąganie XPath za pomocą Google Chrome

5.4 Implementacja

Widok główny aplikacji, na który są renderowane pozostałe widoki jest zdefiniowany w pliku _Layout.cshtml. W kodzie HTML określony jest wygląd paska nawigacyjnego, miejsce w którym będą się ładować poszczególne widoki (@RenderBody()) oraz pliki ze skryptami javascript i css.

W projekcie są zaimplementowane dwa widoki aplikacji. Pierwszym z nich jest „Home”, który wyświetla się po przejściu pod adres URL strony. Jest to zapewnione przez metodę RegisterRoutes klasy RouteConfig. Wewnątrz klasy wywoływana jest metoda MapRoute z klasy RouteCollection, która mapuje konkretne adresy URL i ustawia dla nich domyślne wartości. W tym przypadku adresy mają postać "{controller}/{action}/{id}". Oznacza to, że pierwsza wartość po adresie serwera określa uruchamiany kontroler, następna wywoływaną akcję a ostatnia id zasobu. Domyślna wartość oznacza co będzie przekazywane do serwera jeśli któraś z nich nie zostanie podana przez użytkownika. W tym przypadku jest to kontroler dla widoku „Home”, akcja „Index”, natomiast element id został oznaczony jako opcjonalny.

Struktura kodu HTML widoku HOME jest bardzo prosta, ponieważ składa się z jednej formy. Zawiera ona pole tekstowe, w której użytkownik wpisuje szukanego autora, oraz pole typu „input” oznaczające przycisk „Search”. Klasa HomeController, która reprezentuje kontroler dla widoku strony startowej zawiera tylko jedną metodę Index będącą domyślną funkcją dla danego kontrolera. Po wybraniu przez użytkownika przycisku wyszukiwania następuje przekierowanie do kolejnego widoku „Articles”

Widok „Articles”, który definiuje widok strony z listą znalezionych publikacji nie jest również bardzo rozbudowany. Przy pierwszym załadowaniu wyświetla się napis „Loading...” oraz animacja kręcącego się kółka, symbolizująca ładowanie wyników. Zaraz po wyświetleniu tej strony wywoływana jest w kontrolerze funkcja „ArticlesLoading”.

Klasa kontrolera ArticlesController zawiera trzy funkcje. Domyślna metoda Index zwraca jedynie widok Articles. Dużo bardziej złożoną funkcją jest ArticlesLoading, która wykonuje się zaraz po załadowaniu strony. Na początku działania metody sprawdzane jest czy w ogóle został podany autor do wyszukania. Jeśli nie, zwracany jest pusty widok, który poinformuje użytkownika że nie znaleziono żadnych informacji. W przeciwnym wypadku funkcja dalej się wykonuje i sprawdza warunek czy podane zapytanie różni się od obecnie wyświetlanego (informacja o tym jest przechowywana w sesji użytkownika). Jeżeli nie to nie ma potrzeby pobierania nowych danych i można przejść do operacji na obecnie przechowywanej w pamięci liście artykułów. Są nimi kolejno:

1. Utworzenie obiektu, który będzie modelem dla widoku artykułów.
2. Utworzenie odpowiednich obiektów przechowywanych w sesji.
3. Posortowanie odpowiednio listy.
4. Jeżeli lista jest posortowana malejąco według liczby cytowań (co jest domyślną wartością) zaznaczana jest odpowiednio pierwsza pozycja nie zaliczająca się do indeksu Hirscha, przed którą wyświetlona zostanie pogrubiona linia.
5. Ustawienie wartości paginacji listy (50 elementów na stronie oraz wyświetlenie pierwszego z nich).

Po zakończeniu tych działań zwracany jest nowy widok znalezionych elementów. Nazywa się `_ArticlesList` i jest widokiem częściowym (ang. Partial View). Oznacza to, że nie jest samodzielnym widokiem, posiadającym kontroler oraz wywoływanym przez odpowiedni adres, lecz może być wyświetlany jako część innych stron. W tym projekcie wykorzystanie widoku częściowego do wyświetlania listy publikacji i informacji o nich zostało wymuszone poprzez potrzebę wyświetlenia animacji informującej o pobieraniu danych. Po skończeniu ładowania rekordów animacja jest zastąpiona przez informację o publikacjach szukanego autora.

`_ArticlesList` jest najbardziej rozbudowanym widokiem w aplikacji. Na samej górze strony wyświetlany jest tytuł wraz z wyszukiwanym zapytaniem. Następnie za pomocą znaczników „@” wykorzystywana jest składnia języka C#. Na samym początku wywoływana zostaje funkcja warunkowa, która sprawdza czy przekazany obiekt modelu zawiera listę artykułów. Jeżeli nie, prezentowana jest stosowna informacja. Gdy zostały pobrane publikacje następuje generowanie tabeli. Znowu wykorzystywana jest składnia C#, tym razem pętla „for”. W kolejnych kolumnach wyświetlana jest liczba porządkowa, informacje o publikacji, rok wydania oraz liczba cytowań. Przy iterowaniu po elementach listy sprawdzany jest jeszcze warunek, czy dany element jest oznaczony jako pierwszy nie wliczający się do h-indeksu. Jeśli tak, górna krawędź danego wiersza zostaje pogrubiona.

Ponad tabelą generowany jest inny widok częściowy nazwany `_Statistics`, który jest odpowiedzialny za odpowiednie wyświetlanie obliczonych statystyk. Zawiera również prosty kod HTML, w którym odpowiednie statystyki wyświetlane są w kolejnych paragrafach.

Pobieranie danych odpowiadających wpisanemu przez użytkownika kluczowi odbywa się w metodzie `GetArticles` klasy `ArticlesController`. W pierwszej kolejności po jej wywołaniu utworzone zostają wymagane obiekty: lista artykułów oraz instancja klasy `ParseHelper` wspomagającej parsowanie kodu HTML do obiektów artykułów. Następnie wywoływana jest statyczna funkcja klasy `WOSWebHelper` o nazwie `ExecuteQuery`, która odpowiada za pobranie stron z serwisu Web Of Science, zwracanych na żądanie wyszukiwania danego autora. W ostatniej fazie funkcja `GetArticles` iteruje po pobranych stronach parsując je do obiektów klasy `ArticleItem` tworząc zwracaną oraz wyświetlaną w dalszej części działania aplikacji listę artykułów.

Najważniejszą klasą w projekcie jest `WOSWebHelper`, zawierająca metody odpowiadające za pobieranie danych ze strony Web Of Science. Jedyna z nich jest funkcja `ExecuteQuery`, z którą komunikuje się kontroler widoku artykułów. W pierwszej części implementacji zawiera ona opcję testowego pobrania danych, które polega na pobraniu stron HTML z lokalnych plików. Wymuszone jest to przez brak możliwości przeglądania serwisu Web Of Science bez wykupionej licencji. Strony HTML zostały wcześniej pobrane na serwerze

uczelnianym i są przechowywane na dysku, w folderze zdefiniowanym w pliku konfiguracyjnym. Tam też za pomocą odpowiedniej flagi można ustawić tryb testowy.

Przed właściwym pobieraniem danych z serwera należy uzyskać identyfikator sesji zwany SID. Wykonuje się to poprzez wysłanie żądania metodą GET na stronę <https://www.webofknowledge.com> i wyciągnięcie z odpowiedzi żądanej wartości. Dzięki tak otrzymanemu SID możemy dalej bez przeszkód komunikować się z serwisem.

Po wykonaniu tych operacji funkcja `ExecuteQuery` przechodzi do pobierania pierwszej strony. Otrzymywana jest ona za pomocą wysłania metodą POST odpowiedniego łańcucha znaków, który zawiera przygotowane zapytanie oraz SID, na stronę https://apps.webofknowledge.com/UA_CitedReferenceSearch.do. Jak widzimy, w URL'u jest zdefiniowany typ pobierania (Cited Reference Search), natomiast symbol UA oznacza przeszukiwanie wszystkich dostępnych baz (opcja All Databases). Wymagane dane do pobrania pierwszej strony wyszukiwania oraz odpowiedni adres został uzyskany przy pomocy narzędzi deweloperskich przeglądarki Google Chrome. Wspomniane żądanie na serwer jest wysyłane aż nie zostanie otrzymana strona z błędem, przy czym automat próbuje przynajmniej dziesięć razy. Jeżeli nie uda się pobrać poprawnych danych wyrzucany jest wyjątek.

Z odpowiedzi serwera można wyciągnąć nie tylko dane publikacji autora, ale również łączną liczbę znalezionych stron wyników wyszukiwania. W następnej kolejności pobierane są kolejne strony z danymi autora. Nie jest wymagane już używanie metody POST i wysyłanie danych w postaci zawiłych ciągów znaków. Wystarczy do podstawowego adresu URL (wspomnianego w poprzednim paragrafie) dodać argument „page” z żadaną wartością. Również w tym przypadku pobieranie każdej strony w razie błędu powtarzane jest 10 razy. Jeżeli nie uda się pobrać konkretnej części danych wyrzucany jest wyjątek, co powoduje przerwanie pracy funkcji. Dzięki temu nie istnieje niebezpieczeństwo wyświetlenia niepełnych danych.

Kolejną istotną klasą projektu jest `ParseHelper`, odpowiadająca za wyciąganie z kodu HTML potrzebnych danych i tworzenie pełnego obiektu publikacji. Zawiera zmienne przechowujące odpowiednie znaczniki XPath, służące do prasowania stron internetowych za pomocą biblioteki `HtmlAgilityPack`. Główną metodą klasy jest `ParseHTML`, przyjmująca jako parametr ciąg znaków reprezentujący kod strony w języku HTML oraz zwracająca listę obiektów klasy `ArticleItem`. W początkowej fazie działania funkcja wyciąga z kodu listy ciągów znaków odpowiadające określonym wartościom (między innymi liście autorów, nazwie publikacji, ilości cytowań). Jest to wykonywane za pomocą pozostałych funkcji klasy. Z powodu różnic pomiędzy strukturami kodu HTML w tabeli dla tytułów, autorów i miejsc publikacji wyciąganie tych danych jest rozdzielone od parsowania pozostałych informacji. Dlatego powstały funkcje `GetAuthors`, `GetPublications`, `GetTitles` oraz `GetElements`.

W przypadku pobierania autorów istnieją różnice w pełnych i skróconych listach. Wiele rekordów nie posiada pierwszej z nich, przez co próba użycia znacznika odpowiadającego

pełnej informacji zwraca błąd i wymaga użycie XPath'a dla skróconych nazw. Gdy wyciąganie danych dla obu znaczników nie zwróci wyników pętla parsowania autorów kończy się, a liczba znalezionych elementów oznacza ilość rekordów na stronie. Dzięki tej wartości automat wie ile wykonać iteracji dla pozostałych funkcji.

Tytuły artykułów, konferencji czy książek są wyciągane tylko za pomocą jednego XPath'a. Często na stronie brakuje tej informacji, więc w tym przypadku do zwracanej listy string'ów należy wprowadzić pusty element.

W przypadku tytułów publikacji wyciąganie informacji jest realizowane przy pomocy dwóch znaczników. Pierwszy z nich odpowiada sytuacji, gdy tytuł jest zawarty w elementach „”, który zawiera również link do publikacji. Drugi znacznik jest wykorzystywany gdy wyświetlony zostanie zwykły tekst.

Pozostałe elementy wyszukanych rekordów są pobierane przy pomocy znaczników określających kolumny kolejnych wierszy. W celu poprawienia czytelności tabeli rekordy wyświetlane są w różnych kolorach: naprzemiennie biały z szarym. Wymaga to użycia dwóch różnych znaczników dla każdej informacji. Na końcu iteracji listy z danymi są łączone w taki sposób, aby odpowiadały wierszom na oryginalnej stronie Web OF Science.

Po skończonym wyciągnięciu danych elementy typu string są łączone w obiekty ArticleItem i zwracane jako wynik działania metody ParseHTML.

Modelem widoku _ArticlesList jest obiekt klasy ArticleModel. Zawiera dwa pola i zarządzające nimi właściwości:

- lista typu IPagedList elementów ArticleItem. Użycie powyższego interfejsu jako typu kolekcji umożliwia rozdzielenie rekordów na strony,
- obiekt typu StatisticsHelper zarządzający statystykami listy publikacji.

Klasa ArticleItem zawiera następujące pola:

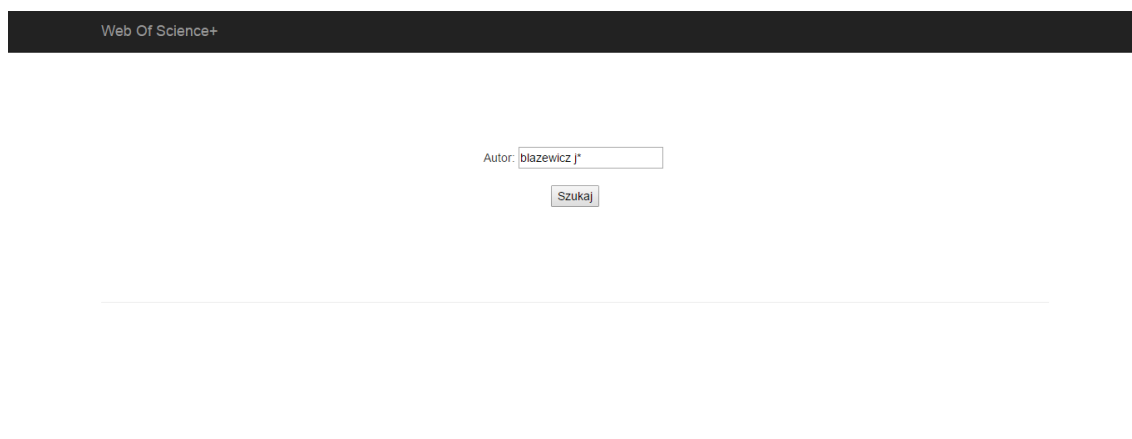
- _Title – tytuł publikacji,
- _Authors – autorzy publikacji,
- _Publication – miejsce publikacji,
- _Volume – wolumin,
- _Issue – wydanie,
- _Pages – strony cytowanego fragmentu,
- _Year – rok wydania,
- _NumOfCitations – liczba cytowań,
- _FirstNotInHIndex – flaga oznaczająca, że dana publikacja jest pierwszą na liście, która nie wlicza się do indeksu Hirscha.

W ArticleItem występuje również właściwość ArticleInfo łącząca podstawowe informacje o publikacji, aby wyświetlić je odpowiednio na stronie.

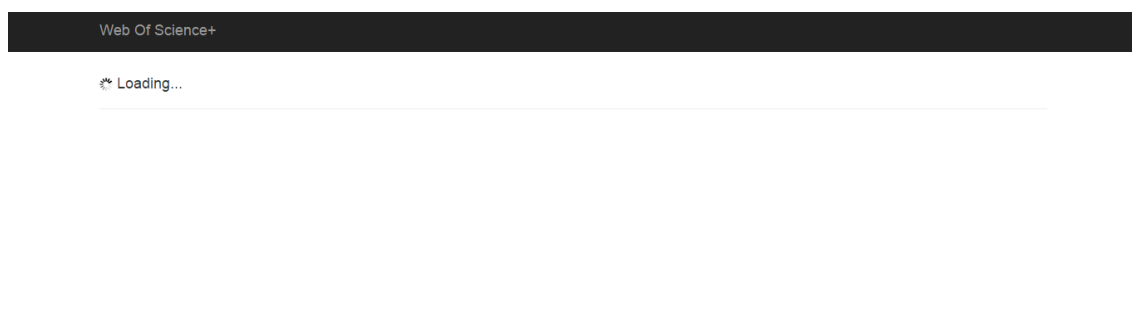
Klasa `StatisticsHelper` przechowuje oraz oblicza wyświetlane statystyki listy artykułów. Są to: liczba publikacji, suma cytowań, średnia liczba cytowań oraz indeks Hirscha. O ile obliczanie podstawowych wartości jest stosunkowo proste to wyliczanie h-indeksu zasługuje na szczególną uwagę. Funkcja za to odpowiadająca iteruje po kolei po liście artykułów posortowanej malejąco po liczbie cytowań. H-index na początku jest równy zero. Otrzymuje wartość równą obecnemu indeksowi pętli powiększonego o jeden jeżeli aktualny artykuł ma liczbę cytowań większą lub równą tej wartości. W przeciwnym wypadku wykonywanie pętli jest zakończone, a obecny h-indeks odpowiada tej statystyce dla całej kolekcji.

5.5 Efekt końcowy

Tak zaimplementowana aplikacja internetowa w pełni spełnia wymagania funkcjonalne. Interfejs użytkownika jest w pełni napisany w języku angielskim. Wyszukanie nazwiska i pierwszej litery imienia autora zwraca zadowalające rezultaty.



Rysunek 20 Wyszukiwanie w aplikacji Web Of Science +



Rysunek 21 Ładowanie wyników wyszukiwania

Znalezione dane prezentowane są w postaci tabeli, którą użytkownik może posortować po liczbie cytowań (domyślnie malejąco) lub roku publikacji oraz w której wyświetlane są pełne nazwy czasopism (o ile znajdują się na oryginalnej stronie). Ponad wynikami wyszukiwania prezentowane jest zestawienie statystyk prac autora: liczba publikacji, suma cytowań, średnia liczba cytowań oraz indeks Hirscha.

Articles List for [blazewicz j*]

Items found: 608

Sum of the Times Cited: 4657

Average Citations per Item: 7.66

h-index: 31

No.	Information	Year	Citations
1	DISCRETE APPLIED MATHEMATICS Title: SCHEDULING SUBJECT TO RESOURCE CONSTRAINTS - CLASSIFICATION AND COMPLEXITY BLAZEWICZ, J. LENSTRA, JK. KAN, AHGR Volume: 5 Issue: 1 Pages: 11 Published: 1983	1983	522
2	Scheduling computer and manufacturing processes Publisher: Springer-Verlag Blazewicz, J.; Ecker, K.H.; Pesch, E.; Schmidt, G.; Weglarz, J. Volume: N/D* Issue: N/D* Pages: N/D* Published: 1996	1996	207
3	EUROPEAN JOURNAL OF OPERATIONAL RESEARCH Title: The job shop scheduling problem: Conventional and new solution techniques Blazewicz, J. Domschke, W. Pesch, E. Volume: 93 Issue: 1 Pages: 1 Published: 1996	1996	188
4	International Journal of Flexible Manufacturing Systems Title: Sequencing of parts and robot moves in a robotic cell Sethi, S.P.; Sriskandarajah, C.; Sorger, G.; Blazewicz, J.; Kubiak, W. Volume: 4 Issue: 3-4 Pages: 331 Published: 1992	1992	157
5	IEEE TRANSACTIONS ON COMPUTERS Title: SCHEDULING MULTIPROCESSOR TASKS TO MINIMIZE SCHEDULE LENGTH BLAZEWICZ, J. DRABOWSKI, M. WEGLARZ, J.	1986	141

Rysunek 22 Wyniki wyszukiwania na stronie Web Of Science +

Po ostatniej publikacji wliczającej się do h-indeksu wyświetlana jest pogrubiona linia

30	BMC BIOINFORMATICS Title: ProCKSI: a decision support system for Protein (Structure) Comparison, Knowledge, Similarity and Information Barthel, Daniel; Hirst, Jonathan D.; Blazewicz, Jacek; Burke, Edmund K.; Krasnogor, Natalio Volume: 8 Issue: N/D* Pages: N/D* Published: 2007	2007	31
31	Publisher: Springer-Verlag, New York, USA Blazewicz, J.; Ecker, K. H.; Pesch, E.; Schmidt, G.; Weglarz, J. Volume: N/D* Issue: N/D* Pages: N/D* Published: 2007	2007	31
32	DISCRETE APPLIED MATHEMATICS Title: On some properties of DNA graphs Blazewicz, J. Hertz, A. Kobler, D. de Werra, D. Volume: 98 Issue: 1-2 Pages: 1 Published: 1999	1999	30
33	DISCRETE APPLIED MATHEMATICS Title: Distributed processing of divisible jobs with communication startup costs Blazewicz, J. Drozdowski, M. Volume: 76 Issue: 1-3 Pages: 21 Published: 1997	1997	30

Rysunek 23 Linia oznaczająca ostatnią pozycję wliczającą się do indeksu Hirscha

Wyniki na stronie są grupowane po 50 elementów.

50	ARTIFICIAL INTELLIGENCE IN MEDICINE Title: Application of tabu search strategy for finding low energy structure of protein Blazewicz, J. Lukasiak, P.; Milostan, M. Volume: 35 Issue: 1-2 Pages: 135 Published: 2005	2005	20
----	---	------	----

*No Data

Page 1 of 13

1	2	3	4	5	6	7	8	9	10	...	»	»»
---	---	---	---	---	---	---	---	---	----	-----	---	----

Rysunek 24 Ostatni element strony i linki do następnych stron

Elementy są pobierane ze strony www.webofknowledge.com, dla typu wyszukiwania Cited Reference Search oraz wszystkich dostępnych baz danych. Na poniższych rysunkach widać, że to samo wyszukiwanie na stronie oryginalnej oraz w aplikacji Web Of Science + zwraca tę samą liczbę rekordów.

WEB OF SCIENCE™

Search

Cited Reference Search

Find the articles that cite a person's work.

Step 2: Select cited references and click "Finish Search."

Hint: Look for [cited reference variants](#) (sometimes different pages of the same article are cited or papers are cited incor

CITED REFERENCE INDEX

References: 1 - 50 of 608

▼

Select Page

Select All*

Clear All

Finish Search

Select	Cited Author	Cited Work [SHOW EXPANDED TITLES]	
<input type="checkbox"/>	Adamiak, RW...Blazewicz, J + [Show all authors]	J COMPUT BIOL	2

Rysunek 25 Liczba znalezionych publikacji na stronie Web Of Science (608)

Web Of Science+

Articles List for [blazewicz j*]

Items found: 608

Sum of the Times Cited: 4657

Average Citations per Item: 7.66

h-index: 31

Rysunek 26 Publikacje znalezione na stronie Web Of Science+ (608)

6. Zakończenie

W czasie realizacji projektu magisterskiego udało się w pełni spełnić przedstawione we wstępie wymagania dotyczące projektowanego systemu. Analiza problemu oceniania dorobku pracownika naukowego uwydatniła potrzebę zbierania danych bibliograficznych i istotę bibliometrii. Jest to bardzo ważne z punktu widzenia naukowca jak i zatrudniającej go jednostki organizacyjnej. Badaczowi pomaga w ocenie, które prace wywarły największy wpływ na środowisko, a jednostce na ocenę pracowników i udzielanie awansów zawodowych. Przedstawienie pojęcia bibliometrii pozwoliło zapoznać się z bardziej miarodajnymi wskaźnikami oceny dorobku naukowego niż suma cytowań czy ich średnia. Najpopularniejszym z nich jest szeroko stosowany indeks Hirscha, według którego najczęściej oceniane są publikacje naukowców.

Przegląd istniejących narzędzi do zbierania danych o publikacjach i cytowaniach pozwoliło zapoznać się z ich funkcjonalnością i możliwościami. Pod tym względem większość z nich nie różni się zbyt wiele dlatego najważniejszym kryterium wyboru portalu mającego być podstawą nowego projektu jest baza publikacji. Największą z nich posiada Google Scholar, lecz przedstawia wiele cytowań o małej wartości bibliometrycznej. Byłby on najlepszym fundamentem nowego rozwiązania ale dyskwalifikuje go brak interfejsu do komunikacji z zewnętrzną aplikacją (API). Wybór zatem padł na portal Web Of Science, który posiada dużą bazę publikacji o wysokiej jakości. Udostępnia jednak opcję szukania cytowanych pozycji bibliograficznych (Cited Reference Search), która pozostawia duże pole do udoskonalania i rozszerzenia funkcjonalności.

Zaimplementowana w ramach pracy aplikacja internetowa Web Of Science + umożliwia użytkownikowi szybkie i proste przeszukiwanie wszystkich baz oryginalnego systemu za pomocą typu wyszukiwania Cited Reference Search. Wyniki wyszukiwania przedstawione są w postaci tabeli, którą można sortować według daty publikacji lub ilości cytowań. System oblicza i prezentuje statystyki znalezionych publikacji: ich liczbę, sumę i średnią liczbę cytowań oraz indeks Hirscha. Implementacja systemu umożliwia jego prostą rozbudowę. Zaleca się dodać w przyszłości obsługę lokalnej bazy danych i zaimplementowanie mechanizmu synonimów oraz listy czasopism punktowanych. Aplikacja została udostępniona na serwerze należącym do Politechniki Poznańskiej, a dostęp do niej możliwy jest z wewnętrznej sieci uczelnianej.

7. Dodatki

Dodatek A. Spis zawartości płyty CD dołączonej do pracy

- Praca magisterska w formie elektronicznej
- Kod źródłowy aplikacji Web Of Science+
- Skompilowany system Web Of Science+

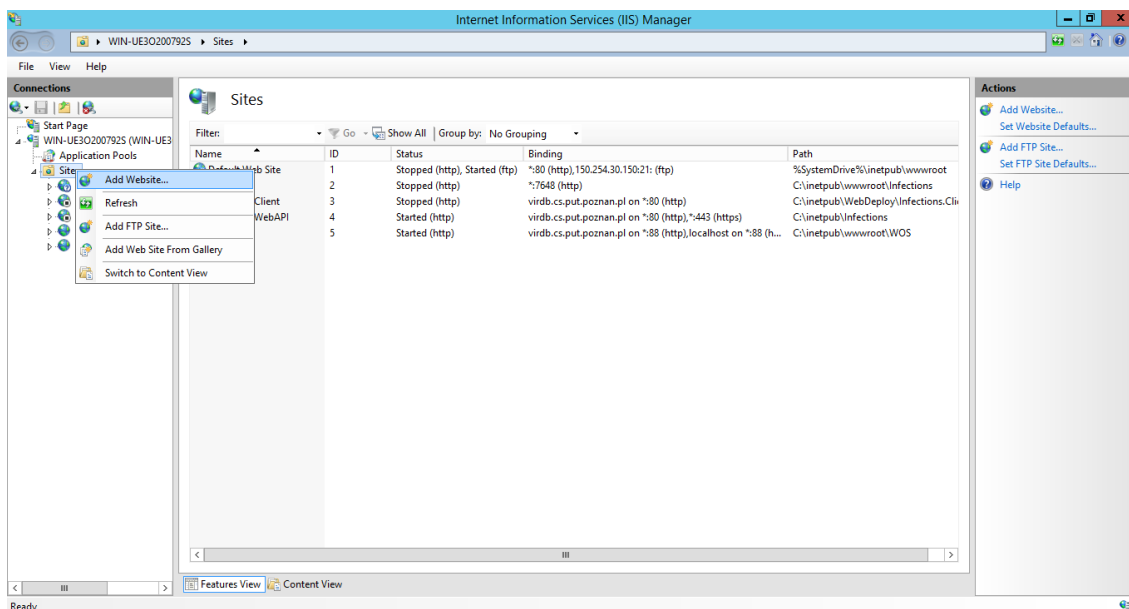
Dodatek B. Instrukcja instalacji i konfiguracji systemu Web of Science+

Wymagania systemu

- serwer z systemem operacyjnym Windows Server 2012 R2 (lub nowszym),
- zainstalowany .NET Framework 4.5.2,
- serwer IIS 6.2.

Instalacja systemu Web Of Science +

1. Uruchomić na serwerze program „Internet Information Services (IIS) Manager”
2. Na liście „Connections” z lewej strony wybrać prawym przyciskiem myszki pozycję „Sites”
3. W rozwiniętym menu kontekstowym wybrać opcję „Add website”



Rysunek 27 Tworzenie nowej strony w IIS

4. W nowym oknie wprowadzić nazwę strony, ścieżkę do folderu z aplikacją (należy wcześniej utworzyć dany folder na dysku) oraz port aplikacji. Po wypełnieniu odpowiednich pól należy kliknąć „OK.”

The screenshot shows the 'Add Website' dialog box in IIS Manager. The 'Site name' field is set to 'WOS'. The 'Application pool' is set to 'WOS'. The 'Physical path' is set to 'C:\inetpub\wwwroot\WOS'. The 'Type' is set to 'http', the 'IP address' is set to 'All Unassigned', and the 'Port' is set to '88'. The 'Start Website immediately' checkbox is checked.

Rysunek 28 Przykładowe dane nowej strony

5. Umieścić w folderze podanym w polu „Physical path” skompilowany system (zamieszczony na płycie CD dołączonej do tej pracy).
6. Aplikacja uruchomi się po wybraniu w przeglądarce internetowej lokalnego adresu IP oraz portu strony

Dodatek C. Kody źródłowe wybranych funkcji aplikacji Web Of Science +

Widok _Layout.cshtml:

```
<!DOCTYPE html>
<html>
<head>
    <meta charset="utf-8" />
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <title>@ViewBag.Title</title>
    @Styles.Render("~/Content/css")
    @Scripts.Render("~/bundles/modernizr")

</head>
<body>
    <div class="navbar navbar-inverse navbar-fixed-top">
        <div class="container">
            <div class="navbar-header">
                @Html.ActionLink("Web Of Science+", "Index", "Home", new { area =
"" }, new { @class = "navbar-brand" })
            </div>
        </div>
    </div>
    <div class="container body-content">
        @RenderBody()
        <hr />
        <footer>
        </footer>
    </div>

    @Scripts.Render("~/bundles/jquery")
    <script src("~/Scripts/Site.js"></script>
    @Scripts.Render("~/bundles/bootstrap")
    @RenderSection("scripts", required: false)
</body>
</html>
```

Klasa RouteConfig

```
public class RouteConfig
{
    public static void RegisterRoutes(RouteCollection routes)
    {
        routes.IgnoreRoute("{resource}.axd/{*pathInfo}");

        routes.MapRoute(
            name: "Default",
            url: "{controller}/{action}/{id}",
            defaults: new { controller = "Home", action = "Index", id =
UrlParameter.Optional }
        );
    }
}
```

Widok Articles:

```
Model WOS_.Models.ArticleModel
@using PagedList.Mvc;
@using PagedList;
@using WOS_.Models;
<link href="/Content/Site.css" rel="stylesheet" type="text/css" />
<link href="/Content/PagedList.css" rel="stylesheet" type="text/css" />

@{
    ViewBag.Title = "Articles";
}

<div class="articles">
    <div class="partialContents" data-url="@Url.Action("articlesloading",
"articles", new { sortOrder = Request.Params["sortOrder"], author =
Request.Params["author"], page = Request.Params["page"] })">
        <br />
        <h4>  Loading...</h4>
    </div>
</div>
```

Widok _ArticlesList:

```
@model WOS_.Models.ArticleModel
@using PagedList.Mvc;
@using WOS_.Helpers
@using WOS_.Models;
<link href="/Content/Site.css" rel="stylesheet" type="text/css" />
<link href="/Content/PagedList.css" rel="stylesheet" type="text/css" />
<h2>Articles List for [@Model.Query]</h2>

@if (Model != null && Model.Articles.Count != 0)
{
    @Html.Partial("_Statistics", Model.Statistics)
    <p>
        <table id="ArticlesList" class="table">
            <tr>
                <th width="5%" style="text-align:center">
                    No.
                </th>
                <th>
                    Information
                </th>
                <th width="10%" style="text-align:center">
                    @Html.ActionLink("Year", "Index", new { sortOrder =
this.Session["DateSortParam"], author = this.Session["CurrentAuthor"] })
                </th>
                <th width="10%" style="text-align:center">
                    @Html.ActionLink("Citations", "Index", new { sortOrder =
this.Session["CitationsSortParam"], author = this.Session["CurrentAuthor"] })
                </th>
            </tr>

            @for(int i =0;i<Model.Articles.Count;i++)
            {
                var item = Model.Articles[i];
                if (item.FirstNotInHIndex)
                {
                    <tr style="border-top:5pt solid black">
```

```

        <td width="5%" style="text-align:center">
            @(i +1+ (Model.Articles.PageNumber - 1) *
Model.Articles.PageSize)
        </td>
        <td>
            <b>@item.Title</b>
            <br />
            @string.Join("; ", item.Authors)
            <br />
            @item.ArticleInfo
        </td>
        <td width="10%" style="text-align:center">
            @item.YearToString()
        </td>
        <td width="10%" style="text-align:center">
            @item.NumOfCitations
        </td>
    </tr>
}
else
{
    <tr>
    <td width="5%" style="text-align:center">
        @(i +1+ (Model.Articles.PageNumber - 1) *
Model.Articles.PageSize)
    </td>
    <td>
        <b>@item.Title</b>
        <br />
        @string.Join("; ", item.Authors)
        <br />
        @item.ArticleInfo
    </td>
    <td width="10%" style="text-align:center">
        @item.YearToString()
    </td>
    <td width="10%" style="text-align:center">
        @item.NumOfCitations
    </td>
    </tr>
}
}
</table>
*No Data
<br />
<br />
<br />
<hr />
    Page @(Model.Articles.PageCount < Model.Articles.PageNumber ? 0 :
Model.Articles.PageNumber) of @Model.Articles.PageCount
    @Html.PagedListPager(Model.Articles, page => Url.Action("Index", new {
page, sortOrder = this.Session["CurrentSort"], author =
this.Session["CurrentAuthor"] }))
    </p>
}
else
{
    <h2>No items found</h2>
}

```

Widok _Statistics:

```
@model WOS_.Helpers.StatisticsHelper
```

```
<div id="statisticsDiv">
  <p><b>Items found:</b> @Model.ItemsFound </p>
  <p><b>Sum of the Times Cited:</b> @Model.SumOfCitations </p>
  <p><b>Average Citations per Item:</b> @Model.AverageNumOfCitations </p>
  <p><b>h-index:</b> @Model.HIndex </p>
</div>
```

Kod funkcji GetArticles:

```
public List<ArticleItem> GetArticles(string author)
{
    List<ArticleItem> articles = new List<ArticleItem>();
    ParseHelper helper = new ParseHelper();
    var foundHtmls = WOSWebHelper.ExecuteQuery(author);
    for (int i = 0; i < foundHtmls.Count; i++)
    {
        var html = foundHtmls[i];
        var a = helper.ParseHTML(html);
        articles.AddRange(a);
        helper.startIndex += helper.count;
    }
    return articles;
}
```

Funkcja ExecuteQuery:

```
public static List<string> ExecuteQuery(string query)
{
    List<string> htmls = new List<string>();
    //Test
    if (Properties.Settings.Default.test)
    {
        foreach (var file in
Directory.GetFiles(Properties.Settings.Default.testFiles))
        {
            htmls.Add(File.ReadAllText(file));
        }
        return htmls;
    }

    HttpResponseMessage response = null;
    int tries = 0;
    string str = "";
    string baseURI = "";
    string sid = "";
    //Getting First Page
    do
    {
        sid = GetSID();
        string postData =
string.Format("fieldCount=3&action=search&product=UA&search_mode=CitedReferenceSe
arch&SID={0}&max_field_count=25&max_field_notice=Notice%3A+You+cannot+add+another
+field.&input_invalid_notice=Search+Error%3A+Please+enter+a+search+term.&exp_noti
ce=Search+Error%3A+Patent+search+term+could+be+found+in+more+than+one+family+%28u
nique+patent+number+required+for+Expand+option%29&input_invalid_notice_limits=+%
3Cbr%2F%3ENote%3A+Fields+displayed+in+scrolling+boxes+must+be+combined+with+at+le
ast+one+other+search+field.&sa_params=UA%7C%7CX2xo9BBvfAf132c3qMF%7Chttps%3A%2F%2
Fapps.webofknowledge.com%3A443%7C%27&formUpdated=true&value%28input1%29={1}&value
%28select1%29=CAU&value%28hidInput1%29=&value%28bool_1_2%29=AND&value%28input2%29
```



```

=&value%28select2%29=CW&value%28hidInput2%29=&value%28bool_2_3%29=AND&value%28input3%29=&value%28select3%29=CY&x=692&y=510&value%28hidInput3%29=&limitStatus=collapsed&ss_lemmatization=On&ss_spellchecking=Suggest&SinceLastVisit.UTC=&SinceLastVisit_DATE=&period=Range+Selection&range=ALL&startYear=1864&endYear=2016&update_back2search_link_param=yes&ssStatus=display%3Anone&ss_query_language=auto&rs_sort_by=PY.D%3BLD.D%3BSO.A%3BVL.D%3BPG.A%3BAU.A", sid, GetProperQuery(query)); //All
databases
        response = PostPage(sid, postData,
@"https://apps.webofknowledge.com/UA_CitedReferenceSearch.do"); //All databases
        str = new StreamReader(response.GetResponseStream()).ReadToEnd();
        baseURI = response.ResponseUri.AbsoluteUri;
        tries++;
    } while (str.Contains("searchErrorMessage") &&
str.Contains("searchErrorMessage") && tries < 10);
    if (tries == 10)
    {
        throw new Exception("Download first page error. Ten tries without
succes");
    }
    int pagesCount = ParseHelper.GetPagesCount(str);
    htmls.Add(str);
    //Getting next pages
    for (int i = 2; i <= pagesCount; i++)
    {
        do
        {
            response = GetPage(string.Format(@"{0}&page={1}", baseURI,
i.ToString()));
            str = new
StreamReader(response.GetResponseStream()).ReadToEnd();
            tries++;
        } while (str.Contains("searchErrorMessage") && tries < 10);
        if (str.Contains("searchErrorMessage") && tries == 10)
        {
            throw new Exception("Download page " + i.ToString() + "
error. Ten tries without succes");
        }
        htmls.Add(str);
    }
    return htmls;
}

```

Funkcja GetSID:

```

static string GetSID()
{
    var response = GetPage(@"https://www.webofknowledge.com?");
    return ParseSID(response.ResponseUri.Query);
}

```

Funkcja ParseSID:

```

static private string ParseSID(string query)
{
    int start = query.IndexOf("SID=") + string.Format("SID=").Length;
    int end = query.Substring(start).IndexOf('&');
    return query.Substring(start, end);
}

```

Funkcja GetPage

```
static private HttpResponseMessage GetPage(string uri)
{
    var request = (HttpRequest)WebRequest.Create(uri);
    request.Method = "GET";
    request.ContentType = "application/x-www-form-urlencoded";
    request.CookieContainer = new CookieContainer();
    Uri target = new Uri("http://www.google.com/");
    request.CookieContainer.Add(new Cookie("CUSTOMER", "ICM Warsaw")
    {
        Domain = target.Host
    });
    request.CookieContainer.Add(new Cookie("E_GROUP_NAME", "Technical
University Poznan")
    {
        Domain = target.Host
    });

    var response = (HttpResponse)request.GetResponse();

    return response;
}
```

Funkcja GetAuthors:

```
List<string> GetAuthors(string html)
{
    List<string> authors = new List<string>();
    HtmlDocument document = new HtmlDocument();
    int authorCount = 0;
    HtmlNode node = null;
    document.LoadHtml(html);
    do
    {
        node =
document.DocumentNode.SelectSingleNode(string.Format(authorsExpXPath, startIndex
+ authorCount));
        if (node == null)
            node =
document.DocumentNode.SelectSingleNode(string.Format(authorsAbrXPath, startIndex
+ authorCount));
        if (node == null)
            break;
        authors.Add(node.InnerText);
        authorCount++;
    } while (node != null);
    count = authorCount;
    return authors;
}
```

Funkcja GetPublications:

```
List<string> GetPublications(string html)
{
    HtmlDocument document = new HtmlDocument();
    document.LoadHtml(html);
    List<string> elems = new List<string>();
    for (int i = 0; i < count; i++)
    {
```

```

        var node =
document.DocumentNode.SelectSingleNode(string.Format(publicationXPath, i +
startIndex));
        if (node == null)
            elems.Add("");
        else
            elems.Add(node.InnerText);
    }
    return elems;
}

```

Funkcja GetTitles:

```

List<string> GetTitles(string html)
{
    HtmlDocument document = new HtmlDocument();
    document.LoadHtml(html);
    List<string> titles = new List<string>();
    for (int i = 0; i < count; i++)
    {
        var node =
document.DocumentNode.SelectSingleNode(string.Format(titleXPath1, i +
startIndex));
        if (node == null)
            node =
document.DocumentNode.SelectSingleNode(string.Format(titleXPath2, i +
startIndex));
        if (node == null)
            titles.Add("");
        else
        {
            titles.Add(string.Join(" ", node.ChildNodes.Select(x =>
x.InnerText)));
        }
    }
    return titles;
}

```

Funkcja GetElements:

```

List<string> GetElements(string html, string whitXPath, string greyXPath)
{
    try
    {
        List<string> white = new List<string>();
        List<string> grey = new List<string>();
        HtmlDocument document = new HtmlDocument();
        document.LoadHtml(html);

        var node = document.DocumentNode.SelectNodes(whitXPath);
        if (node != null && node.Count() != 0)
            white = node.ToList().Select(x => x.InnerText).ToList();

        node = document.DocumentNode.SelectNodes(greyXPath);
        if (node != null && node.Count() != 0)
            grey = node.ToList().Select(x => x.InnerText).ToList();

        var list = ZipLists(white, grey);

        return list;
    }
}

```

```

        catch (Exception e)
        {
            throw new Exception(string.Format("Error: {0}\r\n withXPath {1}
greyXPath {2}", e.Message, whitXPath, greyXPath));
        }
    }

```

Właściwość ArticleInfo:

```

public string ArticleInfo
{
    get
    {
        return string.Format("  Volume: {2} Issue: {3} Pages: {4}",
            _Authors, Publication, Volume.ToString(), Issue.ToString(),
Pages);
    }
}

```

Funkcja CalcHIndex:

```

private int CalcHIndex(List<ArticleItem> articles)
{
    int hindex = 0;
    articles = articles.OrderByDescending(x =>
x.NumOfCitations).ToList();
    for(int i = 0; i < articles.Count; i++)
    {
        if (articles[i].NumOfCitations >= (i + 1))
            hindex = i + 1;
        else
            break;
    }
    return hindex;
}

```

8. Bibliografia

1. Enrique Orduña-Malea, Juan Manuel Ayllón, Alberto Martín-Martín, Emilio Delgado López-Cózar, *About the size of Google Scholar: playing the numbers*, w: *Scientometrics*, 2015, s. 931-949
2. Wikipedia – wolna encyklopedia, *ASP.NET MVC*,
https://pl.wikipedia.org/wiki/ASP.NET_MVC, Data dostępu: 20.10.2016
3. dr hab. Zbigniew Osiński, *Bibliometria metoda analizy i oceny dorobku naukowego historyków najnowszych dziejów Polski*
4. dr Emanuel Kulczycki, *Elementarz: o co chodzi z tymi punktami za publikacje naukowe*, http://ekulczycki.pl/warsztat_badacza/elementarz-o-co-chodzi-z-tymi-punktami-za-publikacje-naukowe/, Data dostępu: 4.10.2016
5. Główny Urząd Statystyczny, *Pojęcia stosowane w statystyce publicznej*,
<http://stat.gov.pl/metainformacje/sloownik-pojec/pojecia-stosowane-w-statystyce-publicznej/15,pojecie.html?pdf=1>, Data dostępu: 4.10.2016
6. dr Emanuel Kulczycki, *Jak obliczyć indeks H?*,
http://ekulczycki.pl/warsztat_badacza/jak-obliczyc-indeks-h/, Data dostępu: 4.10.2016
7. dr Emanuel Kulczycki, *Lista filadelfijska, czyli co?*,
http://ekulczycki.pl/warsztat_badacza/lista-filadelfijska-czyli-co/, Data dostępu: 4.10.2016
8. Microsoft Academic – FAQ,
<https://microsoftacademic.uservoice.com/knowledgebase/articles/838965-microsoft-academic-faq>, Data dostępu: 11.10.2016
9. Wikipedia – wolna encyklopedia, *Microsoft Academic Search*,
https://en.wikipedia.org/wiki/Microsoft_Academic_Search, Data dostępu: 11.10.2016
10. Wikipedia – wolna encyklopedia, *Model-View-Controller*,
<https://pl.wikipedia.org/wiki/Model-View-Controller>, Data dostępu: 20.10.2016
11. *Rozporządzenie Ministra Nauki I Szkolnictwa Wyższego z dnia 27 października 2015 r. w sprawie kryteriów i trybu przyznawania kategorii naukowej jednostkom naukowym*
12. Tomasz Charoński, Monika Grabowska, Mateusz Lango, Jakub Lewiński, *Scholar+. Podręcznik użytkownika*, 2014
13. Tomasz Charoński, Monika Grabowska, Mateusz Lango, Jakub Lewiński, *Scholar+: Rozszerzenie funkcjonalności serwisu Google Scholar*, 2014
14. Wikipedia – wolna encyklopedia, *Tor (sieć anonimowa)*,
[https://pl.wikipedia.org/wiki/Tor_\(sie%C4%87_anonimowa\)](https://pl.wikipedia.org/wiki/Tor_(sie%C4%87_anonimowa)), Data dostępu: 15.10.2016
15. *WEB OF SCIENCE CORE COLLECTION Skrócona Instrukcja*,
<http://wokinfo.com/media/pdf/wos->

corecoll_qrc_pl.pdf?utm_source=false&utm_medium=false&utm_campaign=false,
Data dostępu: 11.10.2016

16. *WEB OF SCIENCE, Web Services Expanded HELP*, <http://ipscience-help.thomsonreuters.com/wosWebServicesExpanded/WebServicesExpandedOverviewGroup/Introduction.html>, Data dostępu: 15.10.2016
17. Wikipedia – wolna encyklopedia, *Web scraping*,
https://en.wikipedia.org/wiki/Web_scraping, Data dostępu: 15.10.2016
18. Wikipedia – wolna encyklopedia, *Wskaźnik Hirscha*,
https://pl.wikipedia.org/wiki/Wska%C5%BAnik_Hirscha, Data dostępu: 4.10.2016
19. Wikipedia – wolna encyklopedia, *Wskaźnik cytowań*,
https://pl.wikipedia.org/wiki/Wska%C5%BAnik_cytowa%C5%84, Data dostępu: 4.10.2016
20. dr Helena Dryzek, *Wskaźnik h, Współczynnik h*,
http://www.mini.pw.edu.pl/~dryzek/www/?S%B3ownik:Wska%BCnik_h%2C_Wsp%F3%B3czynnik_h, Data dostępu: 4.10.2016
21. <https://www.elsevier.com/solutions/scopus/content>, Data dostępu: 11.10.2016
22. <https://scholar.google.pl/intl/pl/scholar/about.html>, Data dostępu: 11.10.2016

9. Spis ilustracji i tabel

Rysunek 1 Główna strona Google Scholar	14
Rysunek 2 Lista wyszukanych elementów	14
Rysunek 3 Przykładowa strona autora	15
Rysunek 4 Widok publikacji	15
Rysunek 5 Basic Search w Web Of Science	17
Rysunek 6 Przykładowy wynik wyszukiwania	17
Rysunek 7 Widok Citation Report	18
Rysunek 8 Lista wyników Cited Reference Search	19
Rysunek 9 Wyszukiwanie w bazie Scopus.....	20
Rysunek 10 Wyszukani autorzy	20
Rysunek 11 Szczegóły autora w bazie Scopus.....	21
Rysunek 12 Wykres dokumentów.....	21
Rysunek 13 Wyszukiwanie w Microsoft Academic	22
Rysunek 14 Lista wyszukanych pozycji w Microsoft Academic.....	23
Rysunek 15 Widok autora	25
Rysunek 16 Widok organizacji.....	26
Rysunek 19 Przeszukiwanie All Databases w trybie Cited Reference Search	33
Rysunek 20 Citation Report.....	34
Rysunek 21 Wyciąganie XPath za pomocą Google Chrome	39
Rysunek 22 Wyszukiwanie w aplikacji Web Of Science +	44
Rysunek 23 Ładowanie wyników wyszukiwania	44
Rysunek 24 Wyniki wyszukiwania na stronie Web Of Science +.....	45
Rysunek 25 Linia oznaczająca ostatnią pozycję wliczającą się do indeksu Hirscha	45
Rysunek 26 Ostatni element strony i linki do następnych stron	45
Rysunek 27 Liczba znalezionych publikacji na stronie Web Of Science (608).....	46
Rysunek 28 Publikacje znalezione na stronie Web Of Science+ (608)	46
Rysunek 29 Tworzenie nowej strony w IIS.....	51
Rysunek 30 Przykładowe dane nowej strony.....	52
 Tabela 1 Porównanie statystyk dla bazy Google Scholar	16
Tabela 2 Porównanie statystyk dla bazy WOS Basic Search	19
Tabela 3 Porównanie statystyk dla bazy WOS Cited Reference Search.....	19
Tabela 4 Porównanie statystyk dla bazy Scopus	22
Tabela 5 Porównanie ilości publikacji poszczególnych autorów w Microsoft Academic	23
Tabela 6 Porównanie danych pochodzących z różnych serwisów dla prof. dr hab. inż. Jacka Błażewicza	26
Tabela 7 Porównanie danych pochodzących z różnych serwisów dla prof. dr hab. inż. Romana Słowińskiego.....	26
Tabela 8 Porównanie danych pochodzących z różnych serwisów dla dr inż. Miłosza Kadzińskiego.....	27