

Aneta Kowalska, Joanna Radzicka
Biblioteka Politechniki Krakowskiej,
Oddział Informacji Naukowej

Programy BibExcel i Pajek w analizach bibliometrycznych

Streszczenie: *Autorki prezentują ogólnodostępne, bezpłatne narzędzia stosowane w analizach bibliometrycznych: program BibExcel pozwalający na zarządzanie danymi, wykonywanie analiz; program Pajek służący do wizualizacji danych, analizy sieci społecznych, do tworzenia map naukowych.*

Słowa kluczowe: *bibliometria, analiza bibliometryczna, BibExcel, Pajek, cytowania, European Summer School for Scientometrics, mapy naukowe, sieci naukowe*

Zainteresowanie auterek programami BibExcel i Pajek zrodziło się w trakcie zajęć European Summer School for Scientometrics [Europejska Letnia Szkoła Bibliometrii] w Wiedniu w 2011 r. BibExcel pozwala na zarządzanie danymi i wykonywanie analiz, których rezultaty można prezentować w formie graficznej np. w programie Pajek.

BibExcel jest to narzędzie, stworzone przez szwedzkiego naukowca Olle Perssona (Inforsk, Umeå Univ, Sweden), służące do analiz bibliometrycznych. BibExcel umożliwia importowanie danych pobranych z baz Scopus, Web of Science (WoS) lub innych źródeł w formie plików tekstowych. Na takich plikach możliwe jest dokonywanie przekształceń i tworzenie różnego typu zestawień, a następnie ich wizualizacja w programach do analizy sieci społecznych, np. Pajek, UCINET, NetDraw. BibExcel umożliwia eksport danych do Excela, jest kompatybilny z systemami Windows oraz Linux. BibExcel jest bezpłatnym programem, dostępnym wraz z materiałami pomocniczymi on-line <http://www8.umu.se/inforsk/Bibexcel/index.html>.

Z kolei Pajek to narzędzie służące do analizy i graficznej prezentacji danych, w tym do tworzenia map naukowych. Autorami programu są Vladimir Batagelj i Andrej Mrvaro ze Słowenii. Pierwsza wersja programu powstała w grudniu 1996 r. Program jest darmowy i można go pobrać z <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>. Strona zawiera opis programu, jego możliwości i sposoby wykorzystania, a także zbiory danych, które można samodzielnie analizować.

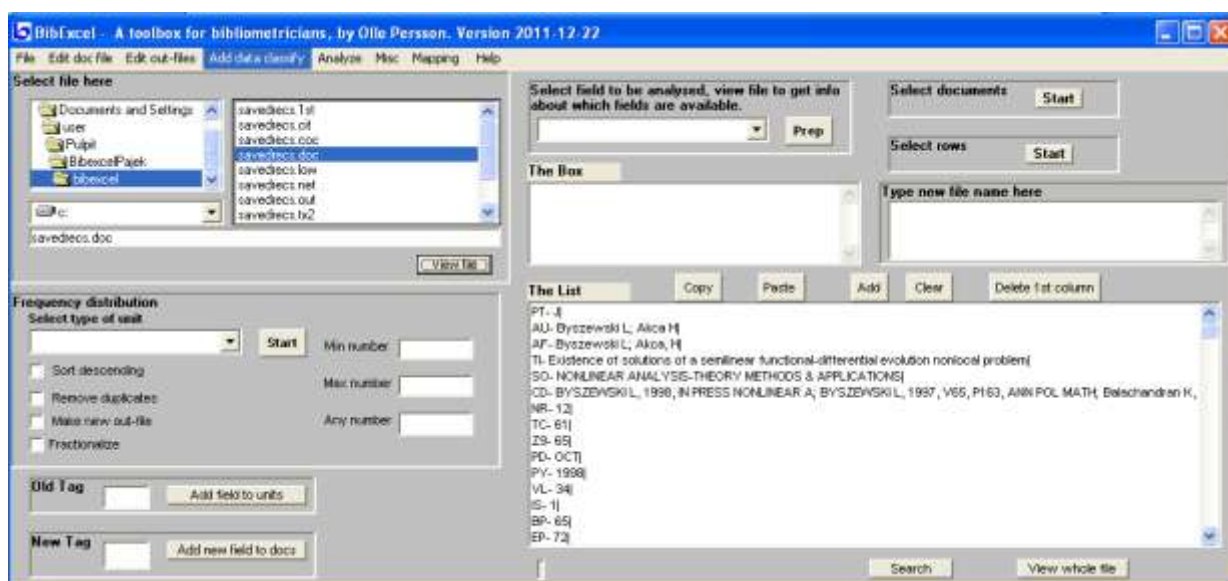
Przygotowanie i eksport danych z bazy WoS i Scopus

Pierwszym krokiem koniecznym do rozpoczęcia pracy w programie BibExcel jest przygotowanie konkretnych danych do analizy. Plik zawierający takie dane powinien być zapisany w formacie tekstowym. W bazie WoS przygotowanie pliku poprzedza zrealizowanie wyszukiwania, na przykład poprzez nazwisko autora lub temat. W rezultacie wyszukiwania należy zaznaczyć wybrane rekordy, np. wykaz publikacji autora i utworzyć listę zaznaczonych dokumentów tzw. "marked list". Po przejściu do "marked list", istnieje możliwość wybrania danych do eksportu (nazwiska autorów, tytuły prac, rok opublikowania i in.). Następnie należy zapisać wybrane rekordy w postaci pliku .txt.

W bazie Scopus etapy przygotowania danych wyglądają podobnie. Różnica pojawia się przy wybieraniu konkretnych rekordów. Po zrealizowaniu wyszukiwania, należy zaznaczyć i wyeksportować wybrane opisy bibliograficzne. Na tym etapie należy zaznaczyć również format pliku, w jakim dane mają być zapisane, w przypadku bazy Scopus jest to format .ris.

Analiza danych w Bibexcel

Pliki o rozszerzeniu .txt. i .ris wymagają konwersji na tzw. "plik wyjściowy" o rozszerzeniu .doc, możliwy do odczytania przez program BibExcel. Ten plik stanowi podstawę do przeprowadzenia kolejnych analiz i przekształceń.



Rys. 1. Okno programu BibExcel.
Źródło: opracowanie własne auterek.

W programie BibExcel poprzez opcję "View file" można zobaczyć pliki. Na rysunku 1. zamieszczony został podgląd opisów bibliograficznych importowanych z baz. W pliku .doc pola rekordu rozdzielone są pionową kreską, elementy pól średnikiem, natomiast koniec rekordu oznaczają dwie pionowe kreski. Na schemacie 1. przedstawiono podgląd rekordu w pliku .doc.

```
PT- J|
AU- Byszewski L; Akca H|
AF- Byszewski L; Akca, H|
TI- Existence of solutions of a semilinear functional-differential evolution nonlocal problem|
SO- NONLINEAR ANALYSIS-THEORY METHODS & APPLICATIONS|
CD- BYSZEWSKI L, 1998, IN PRESS NONLINEAR A; BYSZEWSKI L, 1997, V65, P163, ANN POL MATH; Balachandran K, 1996, V27, P443, INDIAN J PURE AP MAT; Lin YP, 1996, V26, P1023, NONLINEAR ANAL-THEOR; BYSZEWSKI L, 1996, V5, P595, DYNAMIC SYSTEMS APPL; BYSZEWSKI L, 1995, V6, P25, SELECTED PROBLEMS MA;
```

```

BYSZEWSKI L, 1995, V184, MONOGRAPH CRACOW U T; AKCA H,
1995, V2, P179, NONLINEAR TIMES DIG; BYSZEWSKI L, 1991,
V162, P494, J MATH ANAL APPL; BYSZEWSKI L, 1990, V40, P11,
APPL ANAL; WINIARSKA T, 1989, V37, P157, B POLON ACAD SCI
MAT; PAZY A, 1983, SEMIGROUPS LINEAR OP|
NR- 12|
TC- 61|
Z9- 65|
PD- OCT|
PY- 1998|
VL- 34|
IS- 1|
BP- 65|
EP- 72|
DI- 10.1016/S0362-546X(97)00693-7|
JN- NONLINEAR ANALYSIS-THEORY METHODS & APPLICATIONS,
1998, V34, N1, P65-72|
UT- WOS:000075003100005 ER||

```

Schemat. 1. Rekord bibliograficzny w programie BibExcel.
 Źródło: opracowanie własne auterek.

Każde pole rekordu oznaczone jest dwuliterowym skrótem, np.:

- AU — autor (*author*),
- PT — typ publikacji (*publication type*),
- SO — źródło (*source*),
- CD lub CR — dokumenty cytowane, bibliografie załącznikowe (*cited document*),
- PY — rok publikacji (*publication year*),
- TI — tytuł (*title*),
- AB — słowa z abstraktu (*abstract*),
- DE — deskryptory (*descriptor*),
- SC — kategoria tematyczna (*subject category*).

Skróty nazw pól wykorzystuje się jako komendy podczas przeprowadzania konkretnych analiz. BibExcel pozwala analizować zawartość każdego pola, a także jego wybranych elementów.

Tab. 1. Przykładowe typy analiz.

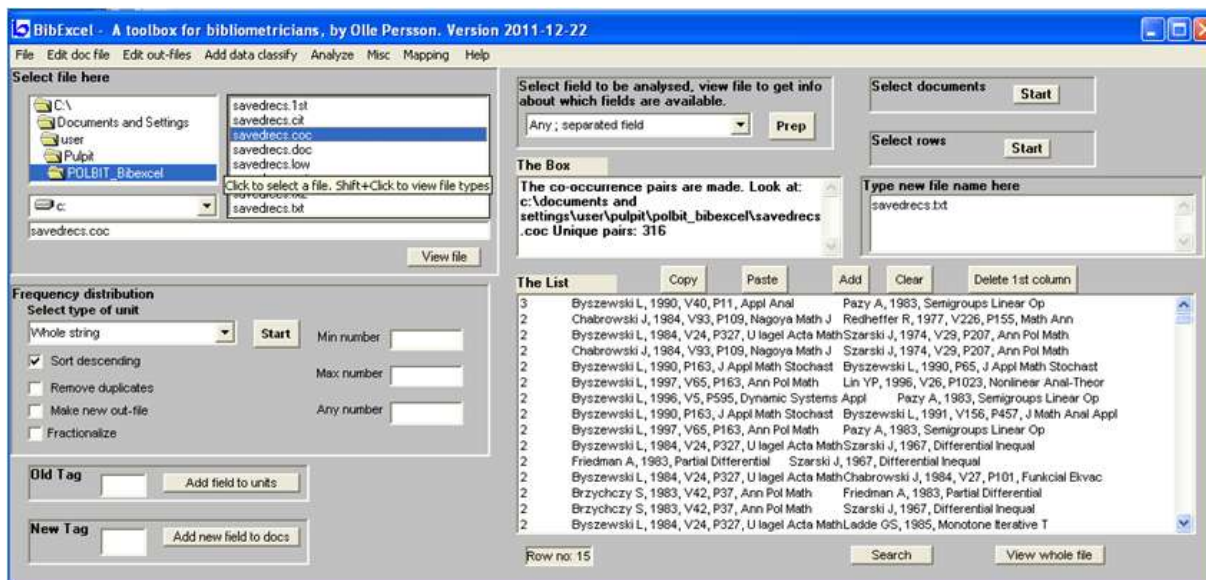
Typ analizy współwystępowania (<i>co-occurrence</i>)	Plik .out musi być oparty na danych z pola	Ścieżka postępowania w menu programu BibExcel
Analiza współcytowań (<i>co-citation analysis</i>)	CR lub CD Analiza całej zawartości pola lub fragmentu, np. autorów cytujących lub publikacji cytujących	Analyze->Co-occurrence->Make pairs via listbox Rezultat: plik .coc
Analiza powiązań bibliograficznych (<i>bibliographic coupling</i>)	CR lub CD Analiza całej zawartości pola lub fragmentu, np. autorów cytujących lub publikacji cytujących	Analyze->Shared units Rezultat: plik .coc
Współautorstwo (<i>co-author analysis</i>)	AU lub AF	Analyze->Co-occurrence->Make pairs via listbox Rezultat: plik .coc

Współwystępowanie słów (<i>co-word analysis</i>)	TI, AB, DE, SC	Analyze->Co-occurrence- >Make pairs via listbox Rezultat: plik .coc
---	----------------	---

Źródło: PERSSON, O. D., DANELL, R., WIBORG SCHNEIDER, J. *How to use Bibexcel for various types of bibliometric analysis. W: Celebrating scholarly communication studies: A Festschrift for Olle Persson at his 60th Birthday, ed. F. Åström, R. Danell, B. Larsen, J. Schneider, p 9–24. Leuven, Belgium: International Society for Scientometrics and Informetrics [on-line]. [Dostęp: 02.03.2012]. Dostępny w World Wide Web: <http://www8.umu.se/inforsk/Bibexcel/ollepersson60.pdf>.*

Program BibExcel pozwala m.in. na:

- przeprowadzanie podstawowych analiz ilościowych oraz częstotliwości występowania poszczególnych słów kluczowych w wybranym wykazie rekordów,
- analizę relacji, współwystępowania (*co-occurrence*), np. autorów, słów z tytułu publikacji,
- analizę powiązań pomiędzy pracami cytowanymi lub cytującymi. Cytowalność można badać, biorąc pod uwagę afiliację autora, miejsca wydania publikacji. Istnieje opcja zawężenia analizy do wybranych lat, co pozwala np. rozważyć wpływ sytuacji politycznej na współpracę autorów,
- redukcję, ograniczenie analizowanego pliku, np. poprzez usunięcie z wykazu elementów o małej liczbie połączeń, relacji; eliminację informacji dublujących się,
- wyliczenie udziału autora przy pisaniu konkretnej pracy,
- wyliczenie indeksu Hirsch'a zarówno dla jednego autora, jak również dla grupy naukowców (z wyszczególnieniem wartości dla każdego z nich).

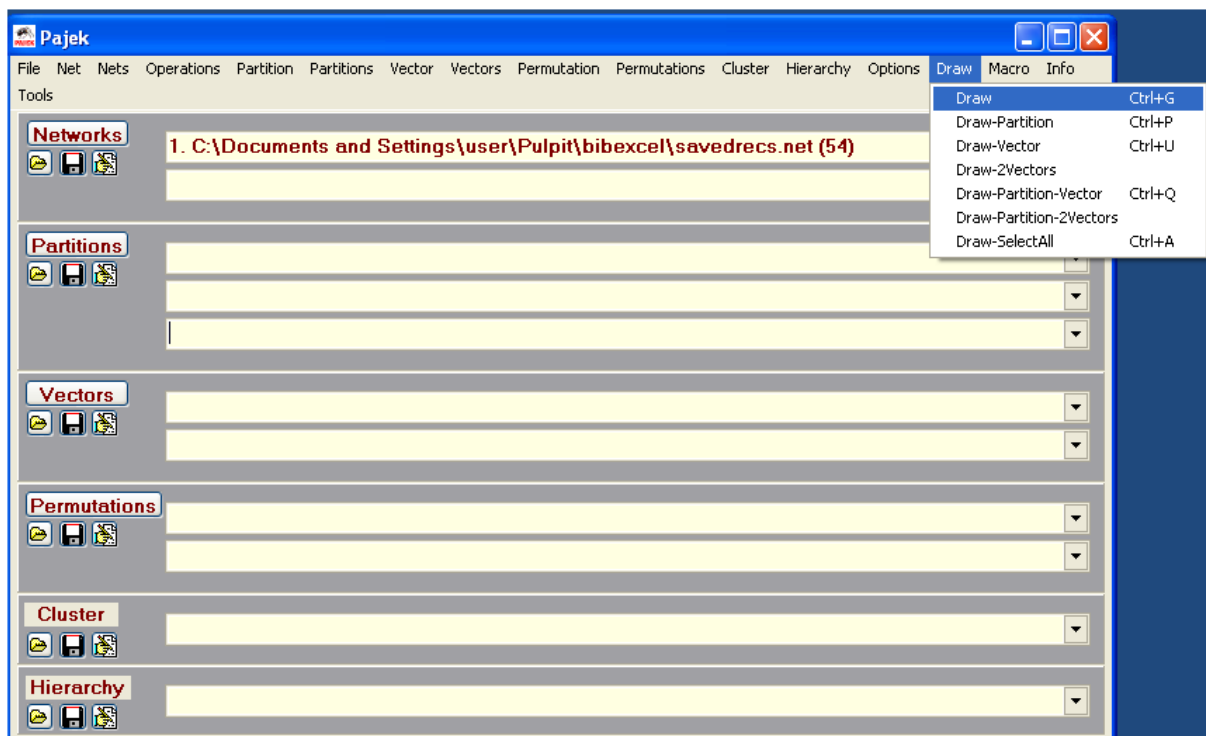


Rys. 2. Dane o współcytowaniach — okno programu BibExcel, plik .coc
 Źródło: opracowanie własne autorek.

BibExcel nie jest programem intuicyjnym, ale po poznaniu podstawowych etapów pracy i komend można samodzielnie przeprowadzać ciekawe i wszechstronne analizy bibliometryczne.

Tworzenie map w programie Pajek

Pliki i dane opracowane w programie BibExcel można przedstawić w formie graficznej, np. za pomocą programu Pajek, który służy do analizy i graficznej prezentacji sieci społecznych. Program nadaje się do analizy stosunkowo dużych sieci. Jest szczególnie przydatny do wstępnych, eksploracyjnych analiz i wizualizacji¹. Główne okno programu Pajek zawiera sześć różnych struktur danych do tworzenia różnego rodzaju wizualizacji (zob. rys. 5.):



Rys. 3. Okno główne w programie Pajek.
 Źródło: opracowanie własne autorek.

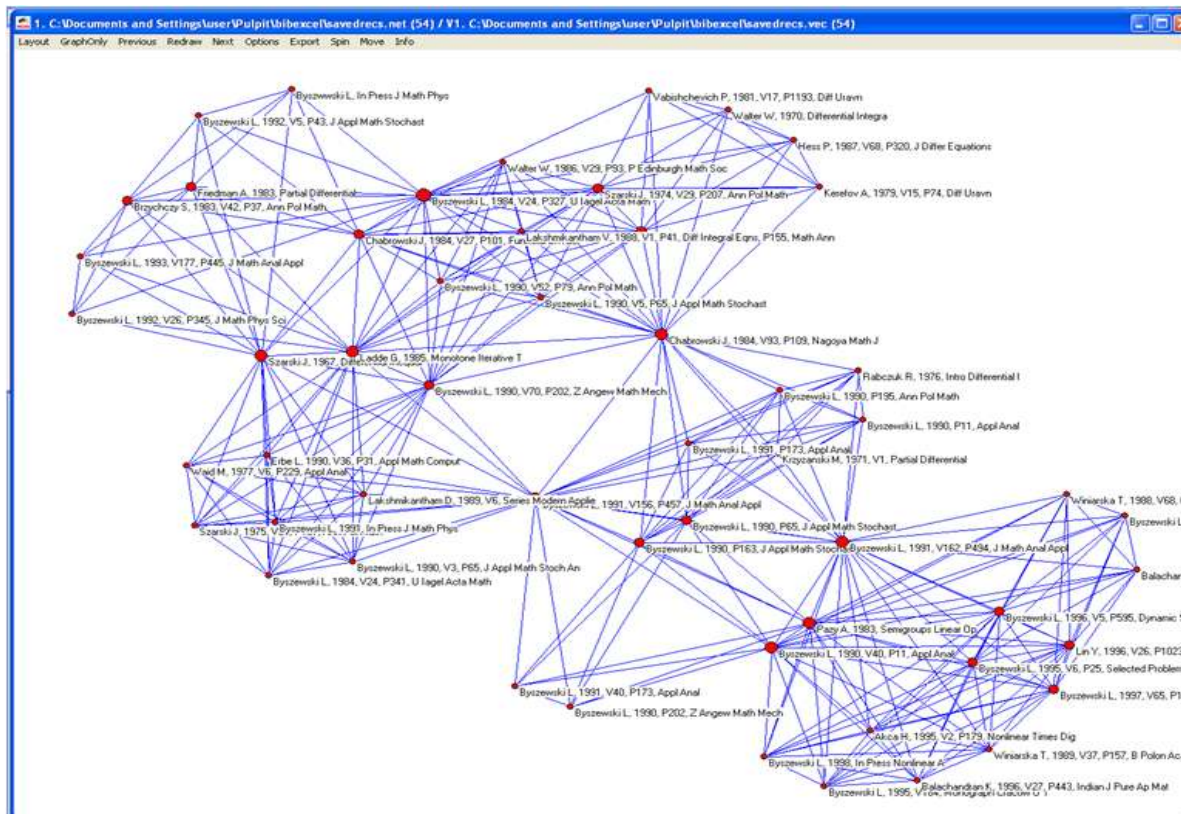
Tab. 2. Rodzaje struktur i danych używanych przez program Pajek.

Struktura	Opis	Rozszerzenie pliku
<i>Network</i>	Sieć — węzły i połączenia	.net
<i>Partition</i>	Węzły i przynależność do grup	.clu
<i>Permutation</i>	Uporządkowanie węzłów	.per
<i>Cluster</i>	Podzbiory węzłów	.cls
<i>Hierarchy</i>	Hierarchicznie uporządkowane klastry i węzły	.hie
<i>Vector</i>	Wektor wartości dla węzłów	.vec

Źródło: BATORSKI, D. Analiza sieci społecznych. Pajek [on-line]. 25 marca 2005. [Dostęp 02.03.2012]. Dostępny w World Wide Web: <http://sna.pl/dbatorski/warsztatPajek.pdf>.

¹ BATORSKI, D. Analiza sieci społecznych. Pajek [on-line]. 25 marca 2005. [Dostęp 02.03.2012]. Dostępny w World Wide Web: <http://sna.pl/dbatorski/warsztatPajek.pdf>.

Dane o sieci zapisuje się w pliku wyjściowym .net, który może być stworzony w zwykłym notatniku, edytorze tekstu lub w programie BibExcel. Plik ten jest podstawą do tworzenia dalszych struktur graficznych. Poza ukazaniem powiązań pomiędzy analizowanymi danymi, można (za pomocą pliku .vec) zaprezentować wielkość danych. Plik ten bierze pod uwagę wartości liczbowe odnoszące się do analizowanych informacji, np. liczbę cytowań autorów (zob. rys. 4.).



Rys. 4. Mapa wektorowa (plik .vec) przedstawiająca współcytowania.
Źródło: opracowanie własne autorek.

Programy BibExcel i Pajek to narzędzia alternatywne względem baz komercyjnych lub narzędzi udostępnianych przez wyszukiwarki naukowe. Mogą one służyć do analiz bibliometrycznych. Istotną cechą programów BibExcel i Pajek jest ich kompatybilność z różnorodnym oprogramowaniem. Pozwalają na wszechstronne badania bibliometryczne, zarządzanie danymi, analizy ilościowe, statystyczne oraz graficzną prezentację dorobku naukowego.

Bibliografia

1. BATAGELI, V., MRVAR, A. Pajek. Program for Analysis and Visualization of Large Networks. Reference. Manual. List of commands with short explanation. version 2.05 [on-line]. 24 September 2011. [Dostęp 02.03.2012]. Dostępny w World Wide Web: <http://vado.fmf.uni-lj.si/pub/networks/pajek/doc/pajekman.pdf>.
2. BATORSKI, D. Analiza sieci społecznych. Pajek [on-line]. 25 marca 2005. [Dostęp 02.03.2012]. Dostępny w World Wide Web: <http://sna.pl/dbatorski/warsztatPajek.pdf>.

3. KOWALSKA, A., RADZICKA, J., Bibexcel i Pajek w analizach bibliometrycznych [on-line]. Warszawa: Politechnika Warszawska, 2012 [Dostęp 10.04.2012]. Dostępny w Internecie: http://www.bg.pw.edu.pl/PolBit/2012/Kowalska_BibexcelPajek.pdf.
4. Więcej: <http://www.bg.pw.edu.pl/PolBit/>
5. KOWALSKA, A., RADZICKA, J. Erasmus i międzynarodowe warsztaty ESSS, Wiedeń 2011 r., Biuletyn EBIB [on-line] 2011, nr 8 (126), Finanse w kulturze [Dostęp 02.03.2012]. Dostępny w World Wide Web: http://www.nowyebib.info/images/stories/numery/126/126_kowalska_radzicka.pdf. ISSN 1507-7187.
6. PERSSON, O. BibExcel [on-line]. [Dostęp 02.03.2012]. Dostępny w World Wide Web: <http://www8.umu.se/inforsk/Bibexcel/>.
7. PERSSON, O. D., DANELL, R., WIBORG SCHNEIDER, J. How to use Bibexcel for various types of bibliometric analysis. W: Celebrating scholarly communication studies: A Festschrift for Olle Persson at his 60th Birthday, ed. F. Åström, R. Danell, B. Larsen, J. Schneider, p 9–24. Leuven, Belgium: International Society for Scientometrics and Informetrics [on-line]. [Dostęp: 02.03.2012]. Dostępny w World Wide Web: <http://www8.umu.se/inforsk/Bibexcel/ollepersson60.pdf>.

Kowalska, A., Radzicka, J. Programy BibExcel i Pajek w analizach bibliometrycznych. W: *Biuletyn EBIB* [online] 2012, nr 3 (130), *Bibliometria w bibliotekach* [Dostęp: 21.04.2012] Dostępny w World Wide Web: http://www.nowyebib.info/images/stories/numery/130/130_kowalska_radzicka.pdf. ISSN 1507-7187.