



WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced  
search

WoS2Pajek

Running

Hints

Analyses

References

# WoS2Pajek networks from Web of Science version 1.4

Vladimir Batagelj

**Manual**

Ljubljana, July 22, 2016 / April 4, 2007



# Outline

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

- 1 Introduction
- 2 Searching
- 3 Advanced search
- 4 WoS2Pajek
- 5 Running
- 6 Hints
- 7 Analyses
- 8 References

**Vladimir Batagelj:** [vladimir.batagelj@fmf.uni-lj.si](mailto:vladimir.batagelj@fmf.uni-lj.si)

**Current version of slides (24. Jul 2016 14:22):** [WoS2Pajek manual](#)



# To be updated . . .

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

Slides beginning with \* are to be updated.

In principle all the options described are still available but in a bit different ways. Some hints:

- to be able to save also CR fields the search has to be limited to Web of Science Core Collection.
- to save the hits they have to be moved to marked and saved using ...



# Searching on the Web of Science

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

Web of Science™ InCites™ Journal Citation Reports® Essential Science Indicators™ EndNote™ Sign In Help English

## WEB OF SCIENCE™

THOMSON REUTERS™

**Search** My Tools Search History Marked List

**Results: 42,424**  
(from Web of Science Core Collection)

You searched for: TOPIC: ("social network") ...More

Create Alert

**Refine Results**

Search within results for...

**Web of Science Categories**

- COMPUTER SCIENCE INFORMATION SYSTEMS (5,414)
- COMPUTER SCIENCE THEORY METHODS (4,183)
- ENGINEERING ELECTRICAL ELECTRONIC (3,230)

Sort by: Publication Date -- newest to oldest Page 1 of 4,243

Select Page Save to EndNote online Add to Marked List

Analyze Results  
Citation Report feature not available. [?]

1. **Measuring Participants' Attitudes Toward Mobile Device Conference Applications in Continuing Medical Education: Validation of an Instrument**  
By: Wittich, Christopher M.; Wang, Amy T.; Fiala, Justin A.; et al.  
JOURNAL OF CONTINUING EDUCATION IN THE HEALTH PROFESSIONS Volume: 36 Issue: 1 Pages: 69-73  
Published: WIN 2016  
POVEZAVE DO GRADIVA View Abstract

Times Cited: 0  
(from Web of Science Core Collection)

Usage Count

2. **Deletion as second death: the moral status of digital remains**  
By: Stokes, Patrick  
ETHICS AND INFORMATION TECHNOLOGY Volume: 17 Issue: 4 Pages: 237-248  
Published: DEC 2016

Times Cited: 0  
(from Web of Science Core Collection)

Usage Count

V. Batagelj

WoS2Pajek 1.4



# Searching on the Web of Science

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced  
search

WoS2Pajek

Running

Hints

Analyses

References

The Web of Science – WoS (ISI/Thomson) allows us to save on a file the records corresponding to our queries.

For example, using Basic search with a query "social network\*"

we get 42424 hits (22. July 2016).

Trying to save them we are informed that we can save at once at most 500 records. We have to save the records by parts on separate files. At the end we concatenate all these files into a single file.



# Saving records from WoS to file

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

- 1 Prepare a list of WoS records to be saved:
  - 1 Select records to be saved either using the **Select Page** box or selecting each record individually. Often it is useful to set the Show option to *50 per page*.
  - 2 Add selected records to the *Marked list* using the button **Add to Marked List**.
- 2 Save records from the Marked list to a file:
  - 1 Select the option *Save to Other File Formats*.
  - 2 In the window **Send to File**
    - 1 Select either the option **All records on page** or specify the range of records to be saved (not more than 500 at once).
    - 2 in **Record Content** select *Full Record and Cited References*.
    - 3 in **File Format** select *Plain Text*.
  - 3 Start saving using the button **Send**. Save the returned file savedrecs.txt to your disk. Close the alert window.



## \* Saving the records

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

Results: **6.936** Show 10 per page Page 1 of 694 Go Sort by: Latest Date

**Output Records**

**Step 1:**

- Selected Records on page
- All records on page
- Records  to

**Step 2:**

- Authors, Title, Source
  - plus Abstract
- Full Record
  - plus Cited Reference

**Step 3:**

[Print](#) [E-mail](#)

[Add to Marked List](#)

[Save to EndNote@Web](#)

[Save to EndNote, RefMan, or other reference software](#)

[Save](#)

At the bottom of the page in the **Output Records** select **Records** and enter the interval bounds *firstRec* to *lastRec* on record numbers that you want to save.

Select Full Record + Cited Reference.

Select also - as Plain Text and click on the **Save** button.



## \* ... Saving the records

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

The screenshot shows the ISI Web of Knowledge interface. At the top, it says "ISI Web of Knowledge™ Take the next step". The main window is titled "Processing Records" and contains the following text:

Please wait while your request is processed.  
(Note: Depending on the number of records, this may take a few moments.)

Product: Web of Science  
Selected action: Save to File  
Processing 500 records:  
10...20...30...40...50...60...70...80...90...100...  
110...120...130...140...150...160...170...180...190...200...  
210...220...230...240...250...260...270...280...290...300...  
310...320...330...340...350...360...370...380...390...400...  
410...420...430...440...450...460...470...480...490...500... Done

Below the progress list are two buttons: "Save" and "Back to Results".

Below the buttons, there is a note: "If the 'Save As' dialog does not appear automatically, then click 'Save' to download the records."

At the bottom left of the main window is the Thomson logo.

Overlaid on the main window is a "Opening savedrecs.txt" dialog box. The dialog contains the following text:

The site has suggested that "savedrecs.txt" be handled as an attachment. It is of type application/octet-stream (Text Document) and located at: <http://pcs.news@knowledge.com>

What should SeaMonkey do with this file?

Open it with the default application (txtfile)

Open it with [ ] Choose...

Save it to disk

Always perform this action when handling files of this type

At the bottom of the dialog are "OK" and "Cancel" buttons.

In a new window the export process starts ...it takes some time ...wait until done. Select Save it to disk and click OK. When the file-chooser appears determine the file on which the records are saved.

Clicking on the **Back to Results** button you return back to the results window.

Repeat these steps until all the records are saved on files.





# \* Using the Advanced Search

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

**ISI Web of Knowledge<sup>SM</sup>** Take the next step

Web of Science | Additional Resources

Search | Cited Reference Search | **Advanced Search** | Search History | Marked List (0)

Web of Science®

**Advanced Search.** Use 2-character tags, Boolean operators, parentheses, and set references to create your query. Results appear in the Search History at the bottom of the page.

Example: TS=(nanclub\* SAME carbon) NOT AU=Smalley RE #1 NOT #2 more examples | view the tutorial

TS=(central\* AND (network\* OR graph))

Search

Field Tags: TS=Topic, TI=Title, AU=Author, GP=Group Author, SO=Publication Name, PY=Year Published, AD=Address, OE=Organization, SG=Suborganization, SA=Street Address, CI=City, PS=Province/State, CU=Country, ZP=Zip/Postal Code

Booleans: AND, OR, NOT, SAME

Current Limits: [Hide Limits and Settings] [Save As My Defaults]

Timespan: All Years (updated 2008-12-20) From 1970 to 2008 (default is all years)

Search History

Set	Results	Combine Sets	Delete Sets
#1	3,199 TS=(central* AND (network* OR graph)) Database=SCI-EXPANDED, SSCI, A&HCI Timespan=All Years	AND OR Combine	Select All Delete

At the computer with access to Web of Science (at Uni-LJ you can use the **IZUM** and select the option **ISI Web of Knowledge (Web of Science)** - na strežniku Thomson Reuters).

Once on the WoS we select the folder **Advanced Search** and enter our query – for example:

TS=(centrali\* AND (network\* OR graph))

If necessary we can set also the time bounds (WoS allows only up to 100000 hits in a query).

We obtain the information about the number of hits at the bottom of the page



# \* Get the list of hits and save selected on file

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

The screenshot shows the Web of Science search results page. At the top, it displays the search criteria: `TS=(central AND (network OR graph))` and the number of results: 3,199. Below this, there are navigation controls for page 1 of 320. A 'Refine Results' sidebar is visible on the left, showing subject areas like 'ENGINEERING, ELECTRICAL & ELECTRONIC (E13)' and 'TELECOMMUNICATIONS (E41)'. The main results list shows a single entry with a blue number '1' next to the title: 'Saturation Throughput Analysis of a Cluster-based Medium Access Control Protocol for Single-hop Ad Hoc Wireless Networks'. Below the title, it lists authors, source, and publication details. At the bottom, there are three steps for saving records: Step 1 (Selected records on page, All records on page, Records 1501 to 2000), Step 2 (Authors, Title, Source, plus Abstract, Full Record, plus Cited Reference), and Step 3 (How do I export to bibliographic management software?). A 'Save to Plain Text' button is highlighted in the Step 3 section.

To get the list of hits we click to their number (blue 3,199 in our case).

At the bottom of this page we can request that some of the hits are saved to the file. For longer lists we have to do this by parts - WoS allows only 500 hits to be saved at once.

To save selected hits we proceed as follows:

- \* **step 1:** determine the range of hits to be saved (1-500, 501-1000, 1001-1500, ...);
- \* **step 2:** select **Full Record** and **plus Cited Reference**;
- \* **step 3:** select **Save to Plain Text**.

Finally we click on the **Save** button.



## \* ... saving

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

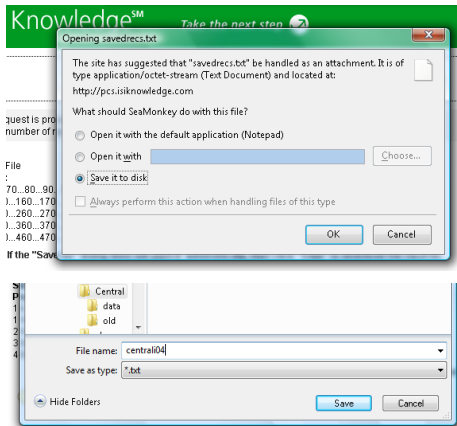
WoS2Pajek

Running

Hints

Analyses

References



A new page **Processing Records** appears. We have to wait until the selected records are processed and written to the file. In the window that appears we select the option **Save to Disk** and click **OK**.

In a new window that appears we select the directory and enter the name of the file on which the selected hits are saved, for example **Centrali004.txt**.

Finally we click on the **Save** button.



# \* ... saving

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

The screenshot shows a green header bar with the text "ISI Web of Knowledge<sup>SM</sup> Take the next step" and a circular refresh icon. Below the header, the title "Processing Records" is centered. The main content area contains the following text: "Please wait while your request is processed. (Note: Depending on the number of records, this may take a few moments.)" followed by "Product: Web of Science" and "Selected action: Save to File". A section titled "Processing 500 records:" lists a sequence of numbers from 10 to 500 in increments of 10, ending with "Done". At the bottom, there are two buttons: "Save" and "Back to Results". The "Save" button is highlighted with a yellow background and has a tooltip that reads: "If the 'Save As' dialog does not appear automatically, then click 'Save' to download the records." The "Back to Results" button is also highlighted with a yellow background and has a tooltip that reads: "When you are done saving the file, click 'Back to Results.'"

To return back to the saving of selected hits we click on **Back to Results**. We repeat the procedure described in this subsection until all the hits are saved.



# \* The list of citing articles

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

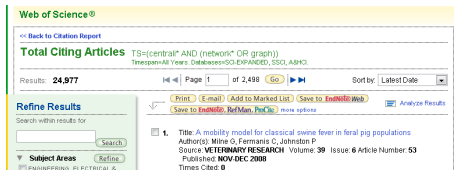
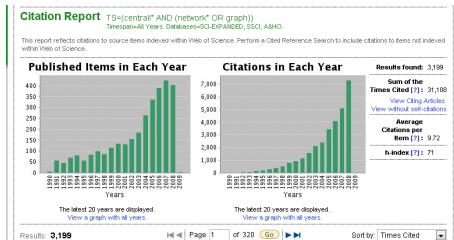
WoS2Pajek

Running

Hints

Analyses

References



We return to the top of the page with list of hits - see the picture in the subsection **Get the list of hits**. In the upper right corner we click on the option **Create Citation Report**. We obtain a new page with histograms. To obtain the list of citing articles we click on the option **View Citing Articles**. To save them we repeat the procedure described in subsection **Save the selected hits to file**.



V. Batagelj

WoS2Pajek 1.4



## \* Additional records

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

Web of Science **Additional Resources**

Search | Cited Reference Search | **Advanced Search** | Search History | Marked List (26)

**Web of Science®**

**Advanced Search.** Use 2-character tags, Boolean operators, parentheses, and set references to create your query. Results appear in the Search History at the bottom of the page.

Example: TS=(nanotub\* SAME carbon) NOT AU=Smalley RE  
#1 NOT #2 [more examples](#) | [view the tutorial](#)

au=(TAKANE Y\*) and py=(1977)

Search

---

**Search History**

Save History / Create Alert | Open Saved History

Combine Sets: AND OR | Delete Sets: Select All Delete

Set	Results		
# 10	1	au=(TAKANE Y*) and py=(1977) Database=SC-EXPANDED, SSC, ABHCI Timespan=All Years	<input type="checkbox"/>
# 9	2	au=(FOWLKES E*) and py=(1993) Database=SC-EXPANDED, SSC, ABHCI Timespan=All Years	<input type="checkbox"/>
# 8	2	au=(CALINSKI T*) and py=(1974) Database=SC-EXPANDED, SSC, ABHCI Timespan=All Years	<input type="checkbox"/>
# 7	3	au=(HARTIGAN J*) and py=(1979) Database=SC-EXPANDED, SSC, ABHCI Timespan=All Years	<input type="checkbox"/>

At WoS we enter the advanced search and for an entry from the list, for example 97

"FELSENST\_J(1985)39:783"

we enter a query

au=(FELSENST\* J\*) and

py=1985

In the list of hits at the bottom of the page click the blue number of hits to obtain the list of their basic descriptions.

Using the information about the volume and the first page, 39 and 783 in our example, identify the corresponding work (if it exists), check the box in front of it and then click the button **Add to Marked List** at the beginning of the list. After addition of the work to the Marked List the red check mark will appear in front of the work (see picture). Repeat the described procedure for other entries.





# \* ... Additional records

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

The screenshot shows the Web of Science search results page. At the top, there are navigation tabs: 'Web of Science', 'Additional Resources', 'Marked List (27)', and 'Search'. The search criteria are 'su=(TAKANE Y\*) and py=(1977)'. The results section shows a list of records, with the first record selected. Below the results, there are options to 'Refine Results' and 'Web of Science Marked Records - 43 Articles'. The 'Step 1. Select the fields to include in the output.' section has several checkboxes checked, including 'Author(s)', 'Title', 'Source', 'Language', 'Abstract\*', 'Cited references\*', 'Document type', 'Subject category', 'Addresses', 'Times cited', 'Keywords', 'Publisher information', 'ISSN', 'Cited reference count', 'Source abbrev.', 'Page count', 'IDS number', and 'Funding information'. The 'Step 2. Select an option.' section has 'Plain Text' selected under 'Format for Print' and 'Plain Text' selected under 'Notes (optional)'. There are also buttons for 'Save to my EndNote Web', 'Save to EndNote, RefMan, ProCite', and 'Save to File'. At the bottom, there is a checkbox for 'Automatically delete selected records from the Marked List after output is complete.'

When the list of hits becomes to long click the **Select All** button in its Delete Sets column and after it the **Delete** button. The list of hits will empty.

To save the works from the Marked List click on Marked List at the top of the page. In the new window select all options in **Step 1** and in **Step 2** select the **Plain Text** option in front of **Save to File** button and click on this button.





# Structure of a WoS record

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

```
PT J
AU KOSMELJ, K
  BATAGELJ, V
TI CROSS-SECTIONAL APPROACH FOR CLUSTERING TIME-VARYING DATA
SO JOURNAL OF CLASSIFICATION
DT Article
CR *UN, 1979, STAT YB
  *UN, 1981, STAT YB
  *UN, 1982, STAT YB
  ANDERBERG MR, 1973, CLUSTER ANAL APPLICA
  BATAGELJ V, 1981, CLUSE CLUSTERING PRO
  BATAGELJ V, 1988, 2ND M YUG SECT CLASS
  BATAGELJ V, 1988, CLASSIFICATION RELAT, P67
  GORDON AD, 1981, CLASSIFICATION
  KOSMELJ K, 1983, REV STAT APPL, V31, P5
  KOSMELJ K, 1986, J MATH SOCIOLOG, V12, P315
TC 7
SN 0176-4268
J9 J CLASSIF
JI J. Classif.
PY 1990
VL 7
IS 1
BP 99
EP 109
SC Mathematics, Interdisciplinary Applications; Psychology, ...
UT ISI:A1990DE57600006
ER
```





# Names of works

WoS2Pajek 1.4

V. Batagelj

The usual *ISI name* of a work (field CR)

LEFKOVITCH LP, 1985, THEOR APPL GENET, V70, P585

has the following structure

AU + ', ' + PY + ', ' + SO[:20] + ', ' + V' + VL + ', ' + P' + BP

All its elements are in upper case.

In WoS the same work can have different ISI names. To improve the precision the program WoS2Pajek supports also *short names* (similar to the names used in HISTCITE output). They have the format:

LastNm[:8] + '\_ ' + FirstNm[0] + '( ' + PY + ' )' + VL + ': ' + BP

For example: LEFKOVIT\_L(1985)70:585

From the last names with prefixes VAN, DE, ... the space is deleted.

Unusual names start with character \* or \$.



# ... Names of works

WoS2Pajek 1.4

In the CR field other forms of ISI names and several errors and inconsistencies can be found:

V. Batagelj

NEWMAN MEJ, 2004, PHYS REV E 2, V69, ARTN 066133  
 PALLA G, 2005, NATURE, V435, P814, DOI 10.1038/nature03607  
 PAPIN JA, 2004, TRENDS BIOCHEM SCI, V29, P641, DOI 10.1016/j.tibs.2004.10.001  
 DOLCINI MM, 2005, J ADOLESCENT HEALTH, V36, UNSP 267.E6-15  
 EVANS JD, 2001, GENOME BIOL, V2, UNSP RESEARCH0001  
 NEWMAN MEJ, 2001, IN PRESS COMPLEX NETUNSP 215239  
 GRANOVET.MS, 1973, AM J SOCIOL, V78, P1360  
 GRANOVETTER M, 1983, SOCIOLOGICAL THEORY, V1, P203  
 BORGATTI SP, 2002, UGINET WINDOWS SOFTW  
 BORGATTI S, 1999, UCINET V USERS GUIDE  
 CANTANZARO M, 2005, PHYS REV E, V71, UNSP 027103  
 CANTAZARO M, 2005, PHYS REV E, V71, UNSP 056104  
 CATANZARO M, 2005, PHYS REV E 2, V71, ARTN 056104  
 BRICKER PD, 1968, OCT M PSYCH SOC ST L : BRICKER

Introduction

Searching

Advanced search

WoS2Pajek

Running



Hints

Analyses

References

We decided to treat in short names the ARTN and UNSP values as BP values. We also remove the DOI parts. There are also irregular names in AU field:

AU BENSON, , C  
 KULHAVY, , W  
 AU SCHONEMA.PH  
 Hansen . T., 1978, THESIS YALE U NEW HA  
 Faradzev . I. A., 1994, INVESTIGATIONS ALGEB, P1

The user can correct the typing errors and nonuniformities on the WoS file.  



# Program WoS2Pajek

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

For converting WoS file into networks in Pajek's format a program **WoS2Pajek** was developed (in Python). It produces the following files:

- citation network: works  $\times$  works;
- authorship (two-mode) network: works  $\times$  authors, for works without complete description only the first author is known;
- keywords (two-mode) network: works  $\times$  keywords, only for works with complete description;
- journals (two-mode) network: works  $\times$  journals, field J9;
- partition of works by the publication year;
- partition of works – complete description (1) / ISI name only (0);
- vector number of pages, PG or EP – BP +1.



# Program WoS2Pajek

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

The keywords are obtained from the fields TI (title), ID, DE and AB (abstract). From the text the **stopwords** are removed and a list of words is produced. The words are lemmatized using **MontyLingua** package.

In future versions additional networks can be derived: works  $\times$  discipline, works  $\times$  countries, . . .

In version 0.7 a GUI support (based on Tkinter) for specifying the program parameters was implemented.

Program WoS2Pajek can be run as an executable program by double-clicking on its icon – see slide 21.

The source code can be executed in different ways using the Python interpreter. See slides 19, 22 and 23.



# Program WoS2Pajek

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

The current version of WoS2Pajek requires 7 parameters to be given by the user:

- MontyLingua directory: path to the directory in which the MontyLingua package is installed (put it also in the PATH env-variable);
- project directory: where the output files are saved;
- WoS file;
- maxnum – estimate of the number of all vertices (number of records + number of cited Works) – 30\* number of records;
- step – prints info about each k\*step record as a trace; step = 0 – no trace.
- use ISI name / short name;
- make a clean WoS file without duplicates;
- boolean list [ DE, ID, TI, AB ] specifying which fields are sources of keywords.



# Program WoS2Pajek– details

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

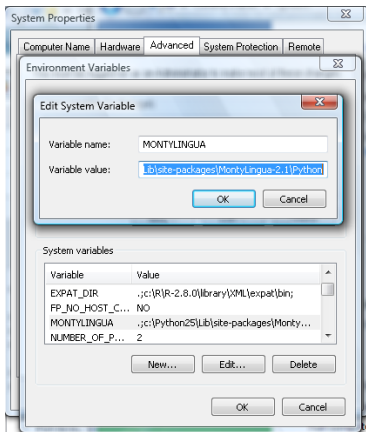
Running

Hints

Analyses

References

To use WoS2Pajek program you need to install at your computer:



- Python, version 2.7
- download WoS2Pajek 1.4 (latest version)
- MontyLingua package
- Copy the MontyLingua package into directory Python27\Lib\site-packages\montylingua-2.1
- add to the environment variable MONTYLINGUA (or PATH) the path to MontyLingua (see the picture): Control Panel/ System/ Advanced System Settings/ Environment Variables/ New



# ... Program WoS2Pajek– details

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

- WoS2Pajek expects in the subdirectory resources (of directory in which it is located) the files StopWords.dat and Pajek.ico;
- run Python and use the commands similar to the following:

```
>>> import sys; wdir = r'c:\users\Batagelj\work\Python\WoS'  
>>> sys.path.append(wdir)  
>>> MLdir = r'c:\Python27\Lib\site-packages\MontyLingua-2.1\Python'  
>>> sys.path.append(MLdir)  
>>> import WoS2Pajek
```

A dialog box will appear in which we specify required parameters and press the RUN button.

WoS2Pajek 0.6 works nicely also on 64-bit computers.



# \* Running WoS2Pajek 0.7 / from Python interpreter

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

```
>>> import sys; wdir = r'c:\users\Batagelj\work\Python\WoS'; sys.path.append(wdir)
>>> MLdir = r'c:\Python25\Lib\site-packages\MontyLingua-2.1\Python'
>>> sys.path.append(MLdir)
>>> import WoS2Pajek
Module WoS2Pajek imported.
```

```
*** WoS2Pajek - 0.7
by V. Batagelj, August 23, 2009 / March 23, 2007
```

```
WoS2Pajek parameters
WoS dir: c:\users\Batagelj\work\Python\WoS
ML dir: c:\Python25\Lib\site-packages\MontyLingua-2.1\Python
Proj dir: C:/Users/Batagelj/work/Python/WoS/batagelj
WoS file: C:/Users/Batagelj/work/Python/WoS/batagelj/batagelj.WoS
MaxNum : 1000
step : 10
ISI name: False
clean : True
keywords: [True, True, False, False]
```

```
***** MontyLingua v.2.1 *****
***** by hugo@media.mit.edu *****
Lemmatiser OK!
Custom Lexicon Found! Now Loading!
Fast Lexicon Found! Now Loading!
Lexicon OK!
LexicalRuleParser OK!
ContextualRuleParser OK!
Commonsense OK!
Semantic Interpreter OK!
Loading Morph Dictionary!
*****
```





# \* ... Running WoS2Pajek 0.7 / from Python interpreter

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

```
*** WoS2Pajek - 0.7
by V. Batagelj, August 23, 2009 / March 23, 2007
started: Mon Aug 24 03:19:29 2009

10 : DOREIAN_P(2000)17:3 - 2009-08-24 03:19:29.614000
20 : BATAGELJ_V(1994)11:93 - 2009-08-24 03:19:30.134000
30 : BATAGELJ_V(1984)52:113 - 2009-08-24 03:19:30.426000
36 : BATAGELJ_V(1975)18:216 - 2009-08-24 03:19:30.640000
>>> End of processing of WoS file

number of works      = 371
number of authors   = 230
number of journals   = 102
number of keywords   = 82
number of records    = 36
number of duplicates = 0
clean WoS data: clean.WoS

*** FILES:
year of publication partition: C:/Users/Batagelj/work/Python/WoS/batagelj\Year.clu
described / cited only partition: C:/Users/Batagelj/work/Python/WoS/batagelj\DC.clu
number of pages vector: C:/Users/Batagelj/work/Python/WoS/batagelj\NP.vec
citation network: C:/Users/Batagelj/work/Python/WoS/batagelj\Cite.net
works X journals network: C:/Users/Batagelj/work/Python/WoS/batagelj\WJ.net
works X keywords network: C:/Users/Batagelj/work/Python/WoS/batagelj\WK.net
works X authors network: C:/Users/Batagelj/work/Python/WoS/batagelj\WA.net
finished: Mon Aug 24 03:19:30 2009
time used: 0:00:01.770000
***

To rerun, type:
    reload(WoS2Pajek)
<module 'WoS2Pajek' from 'c:\users\Batagelj\work\Python\WoS\WoS2Pajek.py'>
>>>
```



# \* Running WoS2Pajek / Python by double-clicking it

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

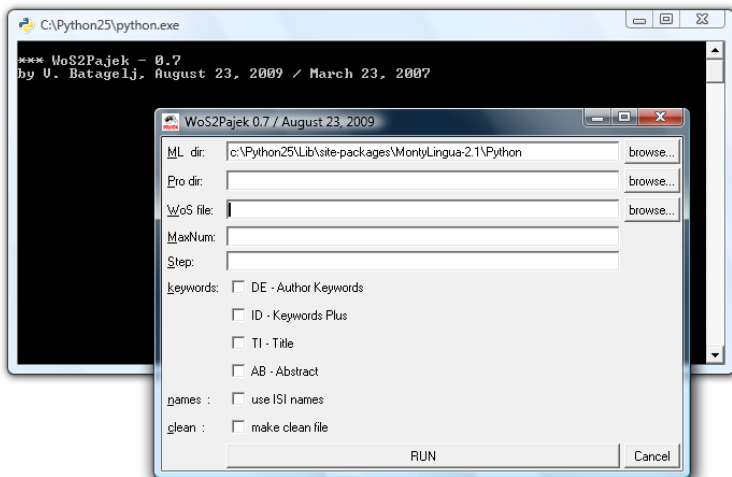
WoS2Pajek

Running

Hints

Analyses

References





# \* Running WoS2Pajek / Python from Dos window

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

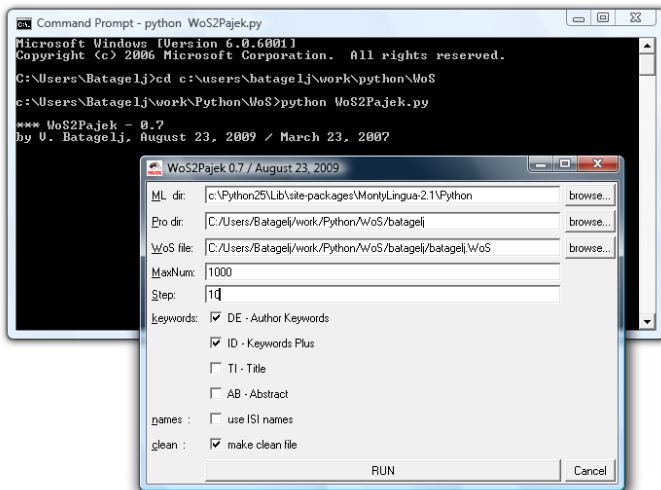
WoS2Pajek

Running

Hints

Analyses

References





# \* Running WoS2Pajek / Python from Dos window using parameters

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

```
Command Prompt
Microsoft Windows [Version 6.0.6001]
Copyright (c) 2006 Microsoft Corporation. All rights reserved.

C:\Users\Batagelj>cd c:\users\batagelj\work\python\WoS

c:\Users\Batagelj\work\Python\WoS>python WoS2Pajek.py c:\Python25\Lib\site-packa
ges\MontyLingua-2.1\Python C:/Users/Batagelj/work/Python/WoS/batagelj C:/Users/B
atagelj/work/Python/WoS/batagelj/batagelj.WoS 10000 10 False True "[True,True,fa
lse,False]"

*** WoS2Pajek - 0.7
by V. Batagelj, August 23, 2009 / March 23, 2007

c:\Python25\Lib\site-packages\MontyLingua-2.1\Python
C:/Users/Batagelj/work/Python/WoS/batagelj
C:/Users/Batagelj/work/Python/WoS/batagelj/batagelj.WoS
10000
10
False
True
[True,True,False,False]

WoS2Pajek parameters
WoS dir:
ML dir: c:\Python25\Lib\site-packages\MontyLingua-2.1\Python
Proj dir: C:/Users/Batagelj/work/Python/WoS/batagelj
WoS file: C:/Users/Batagelj/work/Python/WoS/batagelj/batagelj.WoS
MaxNum : 10000
step : 10
ISI name: False
clean : True
keywords: [True, True, False, False]

***** MontyLingua v.2.1 *****
***** by hugo@media.mit.edu *****
```



# Types on DC file

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

When we combine partial files with saved records from WoS into a single file required by the program WoS2Pajek we can include into this file some additional lines: Comments have the form

**\*\* comment**

Besides this we can specify different types of input records using the lines of the form

**\*T  $n$**

where  $n$  is a type number (1, 2, ...). Since the same record can appear in different parts of the file its class is determined as the set of all corresponding types transformed in integer. For example:  $\{3, 1\} \rightarrow 5$ .



# Analyses

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

The saved records from WoS can still contain some inconsistencies:

- different names for the same person;
- same name for different persons;
- duplicated entries;
- ...

Some of them are detected as results of the analyses. The simplest way to deal with them is to correct them in the saved WoS file and rerun the creation of Pajek's files and analyses.

To improve the quality of the data some tools for detecting (possible) inconsistencies could be developed.

Check (in Pajek) the obtained networks for multiple lines and remove them, if they exist. Remove also the loops from the citation network.



# Preparing the citation network

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

Using on *PR*cite.net the commands

```
Info/Network/General  
Net/Transform/Remove/Loops  
Net/Transform/Remove lines/Single line
```

we get the information about the number of loops and multiple lines, remove loops, and replace multiple lines with single lines. The obtained network we save (Options - Save coordinates [OFF]) to file *PR*citeR.net. For further analysis the citation network has to be acyclic – has no nontrivial strong component. To identify nontrivial strong component and extract them use the commands:

```
Net/Components/Strong [2]  
Operations/Extract from Network/Partition [1-*]  
Operations/Transform/Remove Lines/Between Clusters
```



# ... Preparing the citation network

## Preprint transformation

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

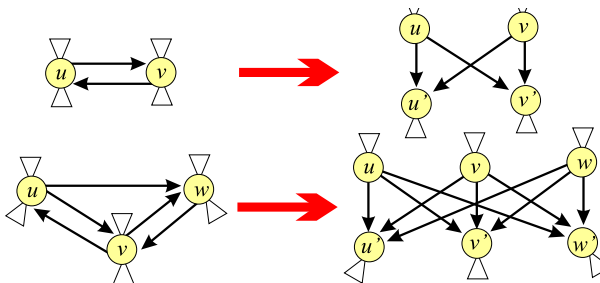
WoS2Pajek

Running

Hints

Analyses

References



To transform the network into acyclic network use the preprint transformation available in Pajek

Network/Acyclic Network/Transform/Preprint Transform





## ... Analyses: network boundary problem

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

Networks obtained from the WoS file using the program WoS2Pajek are in the 'raw' form. We still have to resolve in some way the *network boundary problem*. The first option is to limit the network to the works with complete descriptions – records from the WoS file. We can get a richer network if we decide to include also some referenced (only) works that are referenced often – at least  $k$  times; we delete vertices for which it holds

$$(0 < \text{indeg}(v) < k) \wedge (\text{outdeg}(v) = 0)$$

Net/Partition/Degree/Input

Partition/Binarize [1-(k-1)]

Net/Partition/Degree/Output

Partition/Binarize [0]

[select partition 1]

[select partition 2]

Partitions/Min(V1,V2)

Operations/Extract from Network/Partition [0]



# ...Analyses: collaboration network

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References



Let us denote the citation network with  $C_i$ , and the authorship network with  $WA$ . Then  $C_o = WA^T * WA$  is the *collaboration network*

[Read xyzWA.net]

Net/Transform/2-mode to 1-mode /Columns

Net/Components/Weak [2]

Operations/Extract from Network /Partition [1-\*]

Net/Transform/Remove/Loops

and  $C_a = WA^T * C_i * WA$  is a network of citations between authors.

[5]



# ... Analyses: Bibliographic Coupling and Co-Citation

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

In WoS2Pajek the citation relation means  $u\mathbf{C}iv \equiv ucitesv$ . Therefore the *bibliographic coupling* network  $\mathbf{biCo}$  can be determined as  $\mathbf{biCo} = \mathbf{Ci} * \mathbf{Ci}^T$ .

[Read xyzCite.net]

Net/Transform/1-mode to 2-mode

Net/Transform/2-mode to 1-mode/Rows

Net/Components/Weak [2]

Operations/Extract from Network/Partition [1-\*]

and the *co-citation* network  $\mathbf{coCi}$  can be determined as  $\mathbf{coCi} = \mathbf{Ci}^T * \mathbf{Ci}$ .

Since the network can be quite large we first eliminate the only-cited works.

[Read xyzCite.net]

Net/Partitions/Degree/Output

Operations/Extract from Network/Partition [1-\*]

Net/Transform/1-mode to 2-mode

Net/Transform/2-mode to 1-mode/Columns

Net/Components/Weak [2]

Operations/Extract from Network/Partition [1-\*]

In the analysis of the obtained networks the comparability of units could/should be considered [1].



## ... Analyses: other derived networks

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

The weights  $w(a, p)$  in the *author citation* network

$$\mathbf{ACi} = \mathbf{WA}^T * \mathbf{Ci}$$

counts the number of times author  $a$  cited work  $p$ .

[Read xyzWA.net]

Net/Transform/Transpose/2-mode

[Read xyzCite.net]

Nets/Multiply First \* Second

Net/Components/Weak [2]

Operations/Extract from Network/Partition [1-\*]

Let  $b(\mathbf{A})$  denotes the binarized version of  $\mathbf{A}$ . The *author co-citation* network can be obtained as

$$\mathbf{ACo} = b(\mathbf{ACi}) * b(\mathbf{ACi})^T$$



## ...Analyses: temporal network

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References

We can also transform the citation network into temporal network using the partition of works by publication year:

```
[Read xyzCite.net]
[Read xyzYear.clu]
Vector/Create Identity Vector
Vector/Transform/Multiply by [2008]
Vector/Make Partition/by Truncating
[select as partition 1: xyzYear]
[select as partition 2: obtained from vector]
Operations/Transform/Add/Time intervals determined by Partitions
```



# References I

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References



Batagelj V., Mrvar A.: Density based approaches to network analysis – Analysis of Reuters terror news network. Workshop on Link Analysis for Detecting Complex Behavior (LinkKDD2003, Washington, DC, USA) August 27, 2003.

<http://www.cs.cmu.edu/~dunja/LinkKDD2003/papers/Batagelj.pdf>



Batagelj, V., Doreian, P., Ferligoj, A. and Kejžar, N.: Understanding Large Temporal Networks and Spatial Networks: Exploration, Pattern Searching, Visualization and Network Evolution. Wiley Series in Computational and Quantitative Social Science. Wiley, October 2014.



Batagelj, V., Cerinšek, M. (2013). On bibliographic networks. *Scientometrics*, 96(3), 845–864. doi: 10.1007/s11192-012-0940-1



Cerinšek, M., Batagelj, V. (2015). Network analysis of Zentralblatt MATH data. *Scientometrics*, 102(2015)1, 977-1001.



## References II

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References



Batagelj, V., Zaveršnik, M.: *Fast algorithms for determining (generalized) core groups in social networks*. *Advances in Data Analysis and Classification* 5 (2) (2011), 129-145.



De Nooy, W., Mrvar, A., and Batagelj, V.: *Exploratory Social Network Analysis with Pajek; Revised and Expanded Second Edition. Structural Analysis in the Social Sciences*, Cambridge University Press, September 2011.



Garfield E.: HISTCITE. <http://www.histcite.com/>; [HISTCITE/index](#); [Social networks](#)



Kejžar N., Korenjak-Černe, Batagelj V.: *Network Analysis of Works on Clustering and Classification from Web of Science*. Submitted to Proceedings of IFCS'09 (Dresden, Germany, March 2009).  
[http://pajek.imfm.si/lib/exe/fetch.php?media=dl:gfk1\\_305.pdf](http://pajek.imfm.si/lib/exe/fetch.php?media=dl:gfk1_305.pdf)



Kessler, M. M.: *Bibliographic Coupling between Scientific Papers*. *American Documentation*, 14(1963)1, 10-25.



# References III

WoS2Pajek 1.4

V. Batagelj

Introduction

Searching

Advanced search

WoS2Pajek

Running

Hints

Analyses

References



Small H.: Co-citation In Scientific Literature – New Measure Of Relationship Between 2 Documents. Journal Of The American Society For Information Science, 24(1973)4, 265-269.



WoS2Pajek:

<http://vladowiki.fmf.uni-lj.si/doku.php?id=pajek:wos2pajek>



Web of Science – WoS (ISI/Thomson):

<http://portal.isiknowledge.com/portal.cgi>



Python: <http://www.python.org/>



Py2Exe: <http://www.py2exe.org/>



MontyLingua package:

<http://web.media.mit.edu/~hugo/montylingua/>



Zaveršnik, M., Batagelj, V.: Islands. XXIV. International Sunbelt Social Network Conference, Portorož, May 12–16, 2004. <http://vlado.fmf.uni-lj.si/pub/networks/doc/sunbelt/islands.pdf>