Electronic Information Sources, Mendeley — Bibliography and PDFs Manager

Vladimír Ulman, Igor Peterlík, Jan Obdržálek

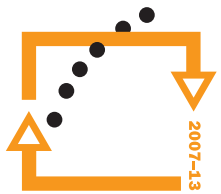
FI MU

DUVOD, October, 2015

european
social fund in the
czech republic

EUROPEAN UNION

MINISTRY OF EDUCATION,
YOUTH AND SPORTS



OP Education
for Competitiveness

INVESTMENTS IN EDUCATION DEVELOPMENT

Impact Factor (IF)

- Defined for *journals*
- The ratio of the number of citations to the previous 2 years of the journal divided by the number of articles in those years
- Essentially the average number of recent citations per article
- Only for journals indexed in *Journal Citation Reports*

Related Measures

- *5-year Impact Factor*
- *Journal Immediacy Index* — the number of citations that year to articles published the same year
- *Journal Citing Half Life* — the median age of the articles that were cited by the articles published in the journal that year
- *Journal Cited Half Life* — the median age of the articles in the journal that were cited by other journals during the year

h-index, after Jorge E. Hirsch (physicist, UCSD)

- For an *individual* scientist
- *h* number of papers that have at least *h* citations each
- Measures *productivity* and *impact* of the published work
- Accessed from *Web of Science* or *Google Scholar*

Original Hirch's suggestion: (for physics!)

12 advancement to tenure (associate professor)

18 full professor

15-20 membership in the US National Academy of Sciences

- Useful only for comparing in the same field
- Grows with *academic age*
- Demonstrated to have high predictive value for National Academy membership or the Nobel Prize

Publication Records: WOS, Scopus

Web of Science = WOS

- http://apps.webofknowledge.com/
- Formerly known as ISI Web of Knowledge
- Operated by the Thomson Reuters
- Provides many tools: IF via JCR, h-index w/o self citations, citation reports etc.
- ResearcherID — a must-have in the academia world in CZ

Scopus

- http://www.scopus.com
- Operated by the Elsevier publishing group

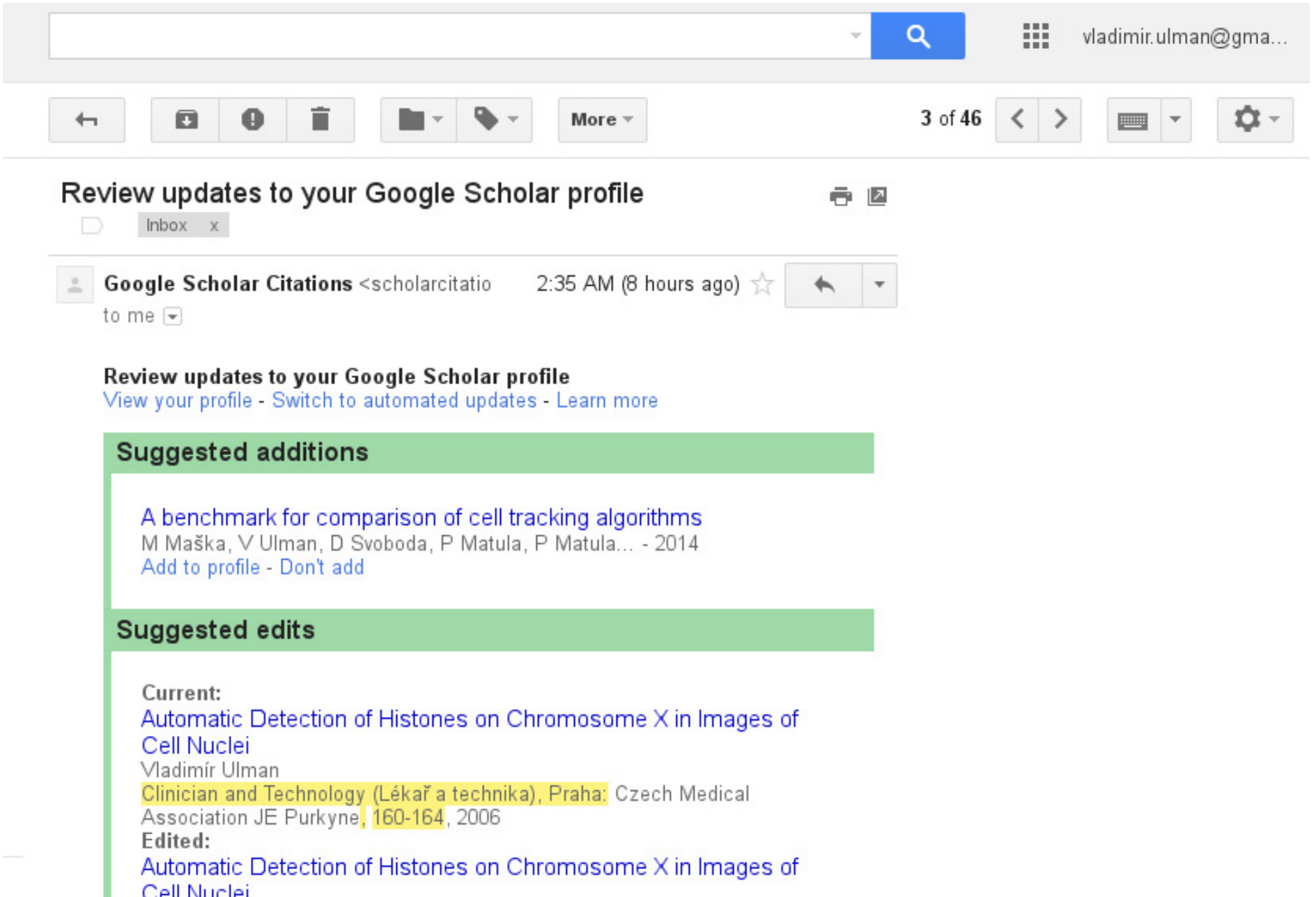
- Also: http://onlinelibrary.wiley.com/
- All services are paid

Publication Records: Google Scholar

- http://scholar.google.cz/
- Google Scholar profile
- Well, Google makes a good use of its data
- It offers free services to users to update and correct links between data

- Service is free of charge

Publication Records: Google Scholar



vladimir.ulman@gma...

More

3 of 46

Review updates to your Google Scholar profile

Inbox x

Google Scholar Citations <scholarcitatio 2:35 AM (8 hours ago)

to me

Review updates to your Google Scholar profile

View your profile - Switch to automated updates - Learn more

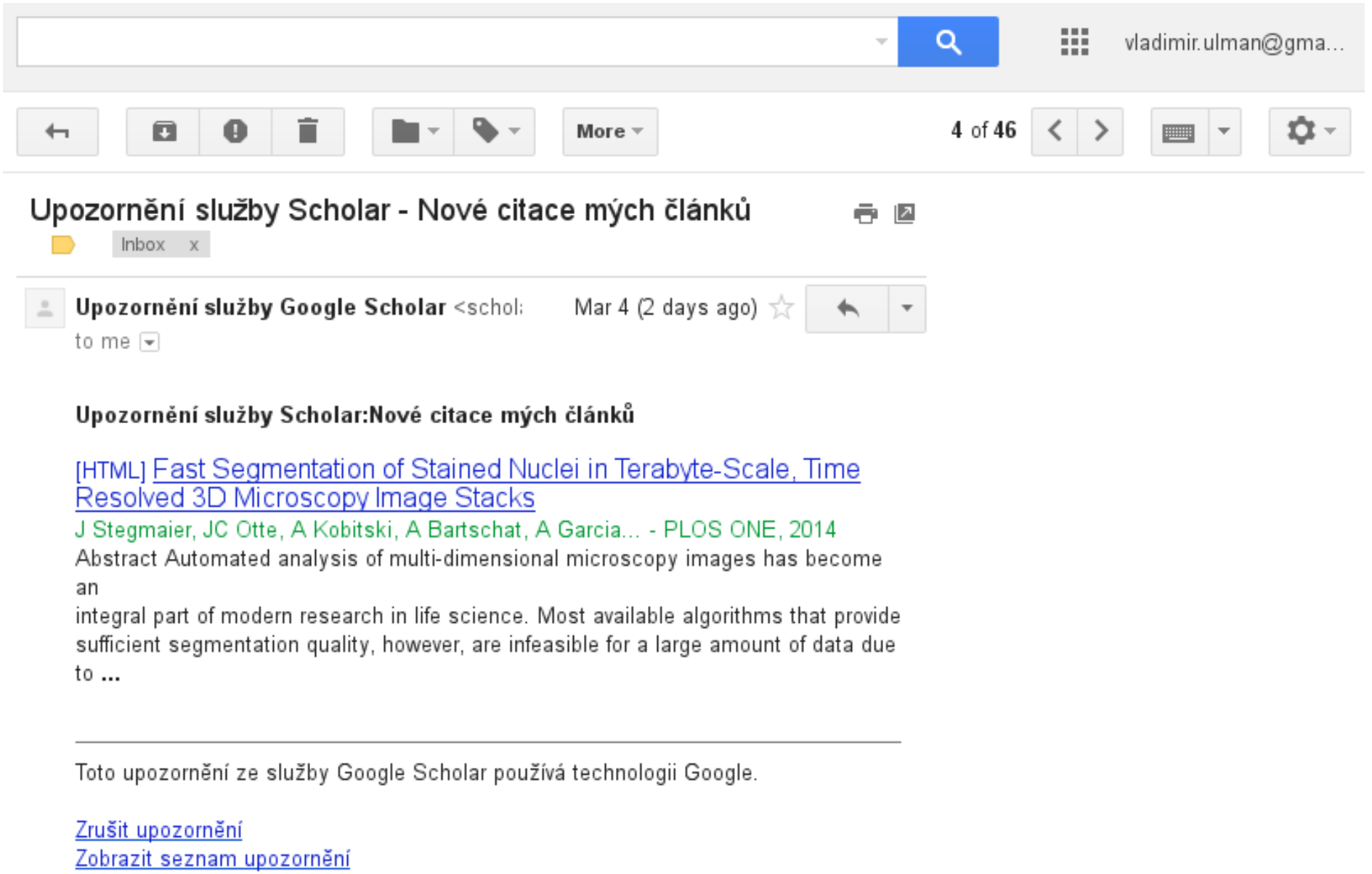
Suggested additions

A benchmark for comparison of cell tracking algorithms
M Maška, V Ulman, D Svoboda, P Matula, P Matula... - 2014
Add to profile - Don't add

Suggested edits

Current:
Automatic Detection of Histones on Chromosome X in Images of Cell Nuclei
Vladimír Ulman
Clinician and Technology (Lékař a technika), Praha: Czech Medical Association JE Purkyne, 160-164, 2006
Edited:
Automatic Detection of Histones on Chromosome X in Images of Cell Nuclei

Publication Records: Google Scholar



Publication Records: DBLP, Microsoft

DBLP

- http://dblp.uni-trier.de/db/
- Operated jointly by Universitat Trier and Schloss Dagstuhl
- Source of good bibliography data, overview of collaborators

Microsoft Academic Search

- http://academic.research.microsoft.com

- Both services are free of charge

Publication Records: Common Issues

- Major problem: *Inconsistent data*
- Completely different numbers
- Errors in data
- Spelling of Czech/Slovak names
- Multiple people with the same name
- Self-citation vs. no self-citations

- GA ČR accepts h-index and citation counts predominantly from WOS and Scopus
- Especially WOS tends to respond to correction requests slowly

Publication Records: Data Sources

- Publication record = Your data + Publisher data
- Supplied by *publisher* to the databases initially, you can correct later on

IS MU

- Building our own list of publication records
- Recently, the list gathers also the publication itself
- Used for generating internal report figures, for *submitting records to RIV*
- RIV checks obtained records against WOS: provide WOS and DOI identifiers

Obtaining Publications

- Publication records databases provide *lists* . . .
- . . . of what a person has published,
- . . . of what has been published in the field.
- But how to obtain the publication itself?
- For most of us: *Google/Google Scholar*
- Directly from publisher (IEEE Xplore, SpringerLink, Science Direct, ACM DIgital Library . . .)
- `http://arXiv.org`, technical reports, dissertation theses, . . .
- From web pages of the lab, the person's homepage, mail request
- Nowadays, databases provide some navigation to the PDF files

Obtaining Publication: DOI

- DOI = Digital Object Identifier
- Example: **10.1000/182** (identifies the DOI Handbook)
- Permanent, resolves to URL
- Resolved through `http://dx.doi.org`
  e.g., `http://dx.doi.org/10.1000/182`

- Not available for old publications

- Alternatively, if available: Use *permalinks* for referring papers
- e.g., `http://www.springerlink.com/content/n62m182vx0227026/`

- People often use Google/Google Scholar for seeking PDFs

Obtaining Publication: Paid Access

- Most of the papers are restricted to download until you pay
- MU has paid and is paying a lot
- Current status: http://ezdroje.muni.cz/
- Access is granted typically based on your IP
- http://vpn.muni.cz
- If this fails, use the portal `ezdroje` above
- LinkSource (formerly SFX): Takes you (somehow) directly to the portal, http://ezdroje.muni.cz/linksource/

Obtaining Publication: arXiv.org

- *Archive* of electronic *preprints*
- Hosted and operated by Cornell University
- Supported by many other institutions
- Guarantees long-term availability
- *Fields:* mathematics, physics, astronomy, computer science, quantitative biology, statistics, and quantitative finance
- *Not* peer reviewed
- Organized by category
- *LaTeX sources* are *compulsory* (if the paper was written in LaTeX)
- Supports *versioning* and *comments*

- followers: `http://viXra.org` and `http://bioRxiv.org`
- relevant: `http://en.wikipedia.org/wiki/List_of_academic_journals_by_preprint_policy`

Obtaining Publication: Good Habits

Keep All PDFs

- Once I manage to download a paper, I'll keep it
- It is not always the case that the paper will be available even the next day (server error, subscription may end)
- Good even for full-text search within the content of the papers

Obtaining Citation Record

- Usually from reference databases, typos can be better detected
- Prefer journal to conference
- Keep full names and titles in your citation records,
  $\rightarrow$ You (or the publ. style) can shorten them anytime
- When referring to software or data:
  $\rightarrow$ Check the web page, cite the tool paper (if it exists)

Managing Bibliography: Storing Papers

How to Organize the PDFs?

- In a nice folder structure with nice names
  $\rightarrow$ works well for most citation managers
- Consider: occasional non-standard access, disaster of the tagging system
- Observation: After some time, the folder structure will change; the filenames will not

How to Organize the Citation Records?

- Human-readable (e.g., .bib for BibTeX or the widespread bibliography format *.ris*) vs. some binary (proprietary) format
- Locally vs. “in a cloud” (e.g., EndNote)
- About citation records: `http://kuk.muni.cz/`

Managing Bibliography: BibTeX

- Traditional, complements *LaTeX*
- Many *frontends*: e.g., JabRef, KBibTeX, . . .
- Inherently desktop-based

- Choose a good naming scheme
- Uppercase letters in titles: *On some {NP}-complete . . .*
- Diacritics and sorting - `{\"a} or {\"{a}} not \"{a}`
- UTF8 *not* working
- `bibtool` — good for managing *bib* files
- List all entries: `\nocite`
- Choose appropriate entry type

Managing Bibliography: Mendeley

- *Mendeley Desktop* - PDF and reference management
- *Mendeley Web* - online social network for researchers
- Platforms (Desktop): Qt based - Windows, Linux, Mac
- Citation data *must* be stored online (free version: 2GB)
- Papers *may* be stored online (you have to set this in folder properties)
- PDFs: *metadata extraction*, *inline comments* (annotations), file organization on disk
- *Bookmarklet* for browsers, working on many websites
- Exports to Word/Libre Office/BibTeX/EndNote
- Multiple computer synchronization (via online space)
- Groups for sharing (pretty limited in the free version: 3 members, up to 100 MB of space)

Managing Bibliography: others

- Overview:
  http://en.wikipedia.org/wiki/Comparison_of_reference_management_software

- BibTeX or Mendeley with LaTeX is dominant in the Comp.Sci.
- EndNote is particularly popular with MS Word in the Biology
- Looking at reference databases, popular tend to be also: BibSonomy, CiteULike, JabRef, RefWorks to name a few
- Zotero: open-source and free, on-line space and sharing
- Integration with Mendeley database
- More paranoid alternatives: store locally, JabRef
- Mind-maps created over the set of PDFs/citations (Docear)