



Open access: why, where and how?

**Publishing Integrity in the Digital Age
Workshop for Journal Editors
University Library in Split, University of Split, Split,
Croatia
6-7 July 2012**

Jadranka Stojanovski
University of Zadar, Ruđer Bošković
Institute

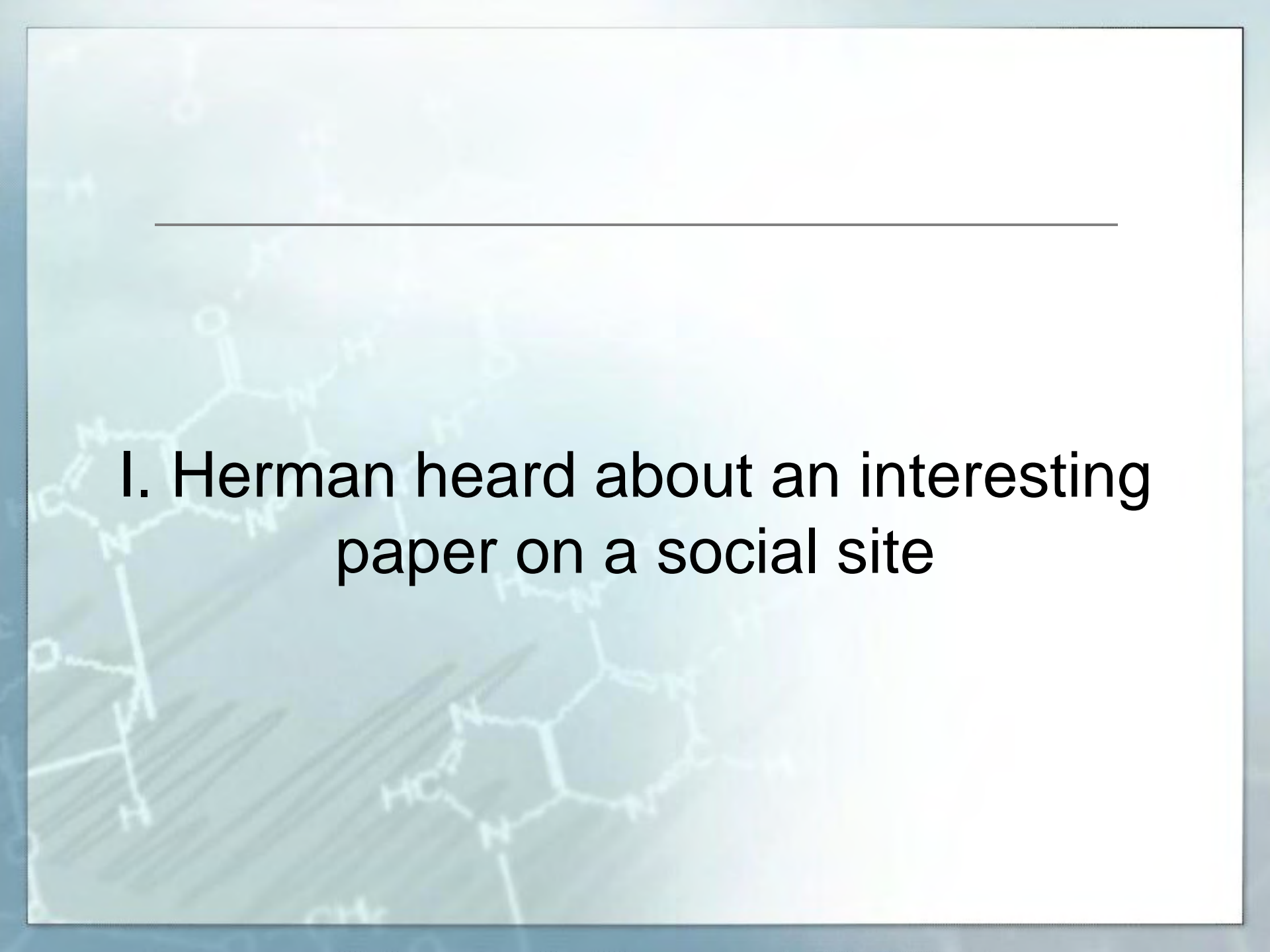
Statements

- no originality
- exaggeration
- no expertize
- not structured well
- not citing properly (but still citing)
 - **maybe boring (I'm a librarian)**



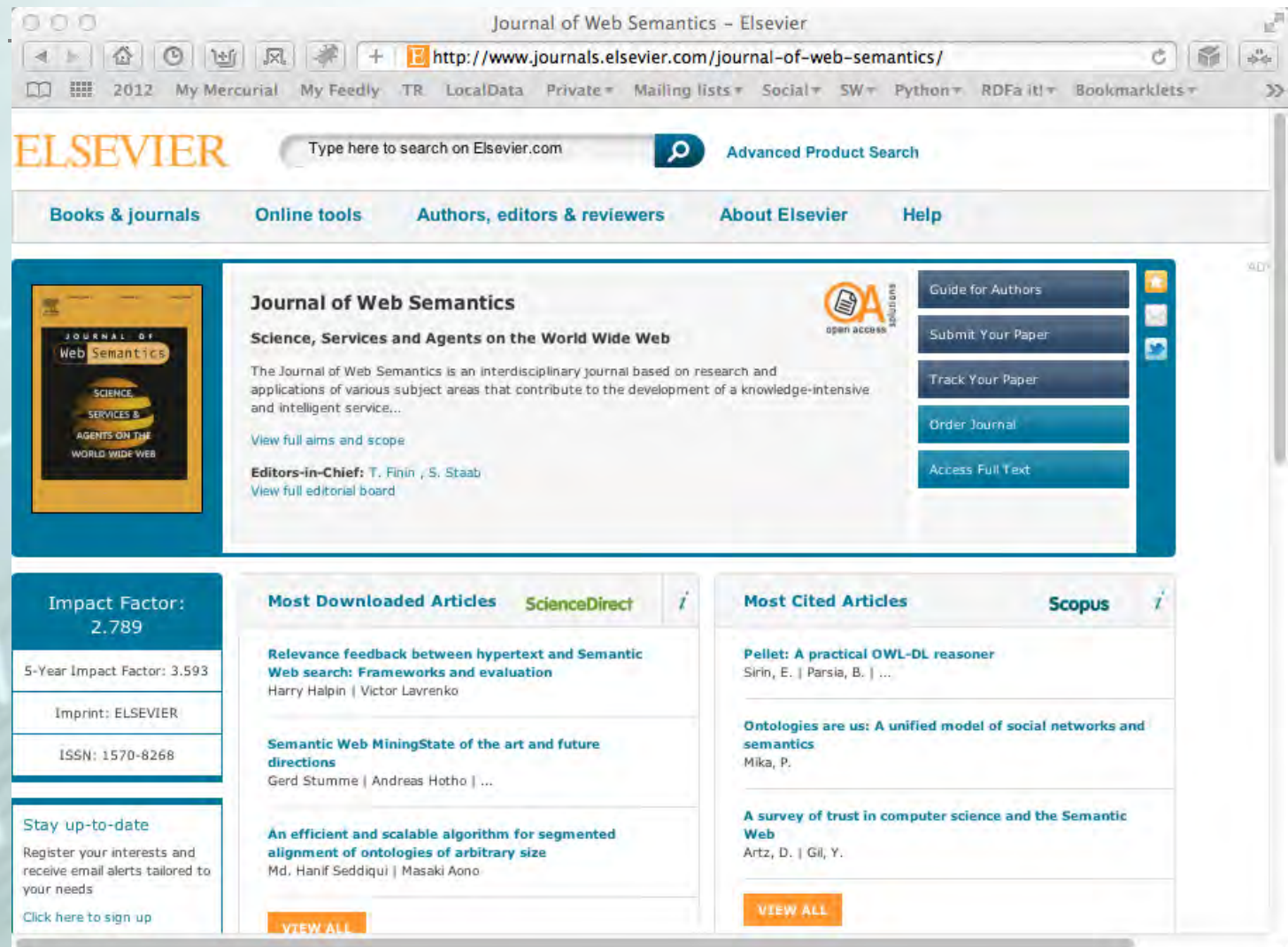
A STORY FROM REAL LIFE

from Ivan Herman: Report on the “Future of Research Communications” Workshop, Dagstuhl, August 15-18, 2012

The background of the slide features a faint, light blue chemical structure, possibly a complex organic molecule or a polymer chain, which is partially obscured by the text and a horizontal line.

I. Herman heard about an interesting
paper on a social site

So he looked up the journal on the Web



The screenshot shows the Elsevier website for the Journal of Web Semantics. The browser address bar displays the URL <http://www.journals.elsevier.com/journal-of-web-semantics/>. The page features the Elsevier logo, a search bar, and navigation links for Books & journals, Online tools, Authors, editors & reviewers, About Elsevier, and Help.

Journal of Web Semantics
Science, Services and Agents on the World Wide Web

The Journal of Web Semantics is an interdisciplinary journal based on research and applications of various subject areas that contribute to the development of a knowledge-intensive and intelligent service...

Editors-in-Chief: T. Finin, S. Staab
[View full editorial board](#)

Impact Factor: 2.789
5-Year Impact Factor: 3.593
Imprint: ELSEVIER
ISSN: 1570-8268

Stay up-to-date
Register your interests and receive email alerts tailored to your needs
[Click here to sign up](#)

Most Downloaded Articles *ScienceDirect*

- Relevance feedback between hypertext and Semantic Web search: Frameworks and evaluation**
Harry Halpin | Victor Lavrenko
- Semantic Web MiningState of the art and future directions**
Gerd Stumme | Andreas Hotho | ...
- An efficient and scalable algorithm for segmented alignment of ontologies of arbitrary size**
Md. Hanif Seddiqui | Masaki Aono

[VIEW ALL](#)

Most Cited Articles *Scopus*

- Pellet: A practical OWL-DL reasoner**
Sirin, E. | Parsia, B. | ...
- Ontologies are us: A unified model of social networks and semantics**
Mika, P.
- A survey of trust in computer science and the Semantic Web**
Artz, D. | Gil, Y.

[VIEW ALL](#)

Open Access
Guide for Authors
Submit Your Paper
Track Your Paper
Order Journal
Access Full Text

Found the paper

Searching and browsing Linked Data with SWSE: The Semantic Services and Agents on the World Wide Web | ScienceDirect.com

http://www.sciencedirect.com/science/article/pii/S1570826811000473


2012 My Mercurial My Feedly TR LocalData Private Mailing lists Social SW Python RDFa it! Bookmarkslets


SciVerse ScienceDirect Hub ScienceDirect Scopus Applications Register Login Go to

Home Browse Search My settings My alerts Shopping cart Help












Back to results Export citation Purchase More options...

Search ScienceDirect

 **Web Semantics: Science, Services and Agents on the World Wide Web**
Volume 9, Issue 4, December 2011, Pages 365–401
JWS special issue on Semantic Search




Searching and browsing Linked Data with SWSE: The Semantic Web Search Engine ☆

Aidan Hogan^a, , , Andreas Harth^b, , , Jürgen Umbrich^a, , , Sheila Kinsella^a, , , Axel Polleres^a, , , Stefan Decker^a, 


^a Digital Enterprise Research Institute, National University of Ireland, Galway, Ireland
^b AIFB, Karlsruhe Institute of Technology, Germany

Available online 22 June 2011.

<http://dx.doi.org/10.1016/j.websem.2011.06.004>, How to Cite or Link Using DOI

 Permissions & Reprints

[View full text](#)



Related articles

- Relevance feedback between hypertext Sem...
Web Semantics: Science, Services and Agents on the World Wide Web
- Sig.ma: Live views on the Web of Data
Web Semantics: Science, Services and Agents on the World Wide Web
- Exploring the Geospatial Semantic Web
Web Semantics: Science, Services and Agents on the World Wide Web
- Chapter 3 - RDF—The basis of the Semantic Web for the Working Ontology
- Scalable and distributed methods for er
Web Semantics: Science, Services and Agents on the World Wide Web

[View more related articles](#)

Related reference work articles
e.g. encyclopedias

- Web Searching
Encyclopedia of Language & Linguistics
- Internet
Encyclopedia of Ecology
- Library Applications
Encyclopedia of Information Systems

[More related reference work articles](#)

However, he was not at his institute...

Searching and browsing Linked Data with SWSE: The Semantic ...vices and Agents on the World Wide Web | ScienceDirect.com

http://www.sciencedirect.com/science/article/pii/S1570826811000473

2012 My Mercurial My Feedly TR LocalData Private Mailing lists Social SW Python RDFa it! Bookmarklets

Back to results Export citation Purchase More options...

Stefan Decker^{a,*}

^a Digital Enterprise Research Institute, National University of Ireland, Galway, Ireland


^b AIFB, Karlsruhe Institute of Technology, Germany

Available online 22 June 2011.

<http://dx.doi.org/10.1016/j.websem.2011.06.004>, How to Cite or Link Using DOI

Permissions & Reprints

[View full text](#)

 [Purchase](#)

Abstract

In this paper, we discuss the architecture and implementation of the Semantic Web Search Engine (SWSE). Following traditional search engine architecture, SWSE consists of crawling, data enhancing, indexing and a user interface for search, browsing and retrieval of information; unlike traditional search engines, SWSE operates over RDF Web data – loosely also known as Linked Data – which implies unique challenges for the system design, architecture, algorithms, implementation and user interface. In particular, many challenges exist in adopting Semantic Web technologies for Web data: the unique challenges of the Web – in terms of scale, unreliability, inconsistency and noise – are largely overlooked by the current Semantic Web standards. Herein, we describe the current SWSE system, initially detailing the architecture and later elaborating upon the function, design, implementation and performance of each individual component. In so doing, we also give an insight into how current Semantic Web standards can be tailored, in a best-effort manner, for use on Web data. Throughout, we offer evaluation and complementary argumentation to support our design choices, and also offer discussion on future directions and open research questions. Later, we

Related reference work articles
e.g. encyclopedias

- Web Searching
Encyclopedia of Language & Linguistics
- Internet
Encyclopedia of Ecology
- Library Applications
Encyclopedia of Information Systems

► More related reference work articles

But... He knew the secret!

About the Preprint Server

http://www.websemanticsjournal.org/index.php/ps/about

Reader

2012 My Mercurial My Feedly TR LocalData Private Mailing lists Social SW Python RDFa it! Bookmarks

Journal of Web Semantics: PREPRINT SERVER
The Preprint Server provides readers with free electronic access to article preprints of the Journal of Web Semantics: Science, Services and Agents on the World Wide Web at Elsevier.

USER
Username
Password
☐ Remember me
Log In

HOME ABOUT LOG IN REGISTER SEARCH CURRENT ARCHIVES ANNOUNCEMENTS

Home > About the Preprint Server

The *Journal of Web Semantics: Preprint Server* provides readers with free electronic access to article preprints of the Journal of Web Semantics: Science, Services and Agents on the World Wide Web at Elsevier.

- » [Aims & Scope](#)
- » [People](#)
- » [Privacy Statement](#)
- » [Submissions](#)
- » [Sponsorship](#)
- » [Site Map](#)
- » [About this Publishing System](#)

AIMS & SCOPE

The Journal of Web Semantics is an interdisciplinary journal based on research and applications of various subject areas that contribute to the development of a knowledge-intensive and intelligent service Web.

These areas include: knowledge technologies, ontology, agents, databases and the semantic grid, obviously disciplines like information retrieval, language technology, human-computer interaction and knowledge discovery are of major relevance as well. All aspects of the Semantic Web development are covered. The publication of large-scale experiments and their analysis is also encouraged to clearly illustrate scenarios and methods that introduce semantics into existing Web interfaces, contents and services. The journal emphasizes the publication of papers that combine theories, methods and experiments from different subject areas in order to deliver innovative semantic methods and applications.

The Journal of Web Semantics addresses various prominent application areas including: e-business, e-community, knowledge management, e-learning, digital libraries and e-sciences.

Readers are also encouraged to visit the Journal of Web Semantics blog, at <http://journalofwebsemantics.blogspot.com/> for more information and related links.

The Journal of Web Semantics includes, but is not limited to, the following major technology areas:

JOURNAL CONTENT
Search
All
Search

Browse
[By Issue](#)
[By Author](#)
[By Title](#)

Search all
[Research Papers](#)
[Survey Papers](#)
[Ontology Papers](#)
[System Papers](#)

JWS BLOG
[JWS special issue: reasoning with...](#)
[Journal of Web Semantics, volume 11...](#)
[JWS special issue on scalability, v10...](#)
[JWS special issue on semantic...](#)
[Top Web technology conferences based...](#)

AUTHOR
[Aims & Scope](#)
[Guide for Authors](#)
[Submit an Article](#)

INFORMATION
For Readers

Go to "http://www.websemanticsjournal.org/index.php/ps/about"

But...He knew the secret!

Searching and Browsing Linked Data with SWSE: the Semantic ...mantics: Science, Services and Agents on the World Wide Web

http://websemanticsjournal.org/index.php/ps/article/view/240

2012 My Mercurial My Feedly TR LocalData Private Mailing lists Social SW Python RDFa it! Bookmarks

Journal of Web Semantics: PREPRINT SERVER

The Preprint Server provides readers with free electronic access to article preprints of the Journal of Web Semantics: Science, Services and Agents on the World Wide Web at Elsevier.

USER

Username

Password

☐ Remember me

Log In

HOME ABOUT LOG IN REGISTER SEARCH CURRENT ARCHIVES ANNOUNCEMENTS

Home > Vol 9, No 4 (2011) > Hogan

SEARCHING AND BROWSING LINKED DATA WITH SWSE: THE SEMANTIC WEB SEARCH ENGINE

Aidan Hogan, Andreas Harth, Juergen Umrh, Sheila Kinsella, Axel Polleres, Stefan Decker

ABSTRACT

Abstract: In this paper, we discuss the architecture and implementation of the Semantic Web Search Engine (SWSE). Following traditional search engine architecture, SWSE consists of crawling, data enhancing, indexing and a user interface for search, browsing and retrieval of information; unlike traditional search engines, SWSE operates over RDF Web data -- loosely also known as Linked Data -- which implies unique challenges for the system design, architecture, algorithms, implementation and user interface. In particular, many challenges exist in adopting Semantic Web technologies for Web data: the unique challenges of the Web -- in terms of scale, unreliability, inconsistency and noise -- are largely overlooked by the current Semantic Web standards. Herein, we describe the current SWSE system, initially detailing the architecture and later elaborating upon the function, design, implementation and performance of each individual component. In so doing, we also give an insight into how current Semantic Web standards can be tailored, in a besteffort manner, for use on Web data. Throughout, we offer evaluation and complementary argumentation to support our design choices, and also offer discussion on future directions and open research questions. Later, we also provide candid discussion relating to the difficulties currently faced in bringing such a search engine into the mainstream, and lessons learnt from roughly six years working on the Semantic Web Search Engine project.

Full Text: [PDF](#)

ARTICLE TOOLS

[Print version](#)

[Indexing metadata](#)

[How to cite item](#)

[Email this article](#)

(Login required)

JOURNAL CONTENT

Search

All

Search

Browse

[By Issue](#)

[By Author](#)

[By Title](#)

Search all

[Research Papers](#)

[Survey Papers](#)

[Ontology Papers](#)

[System Papers](#)

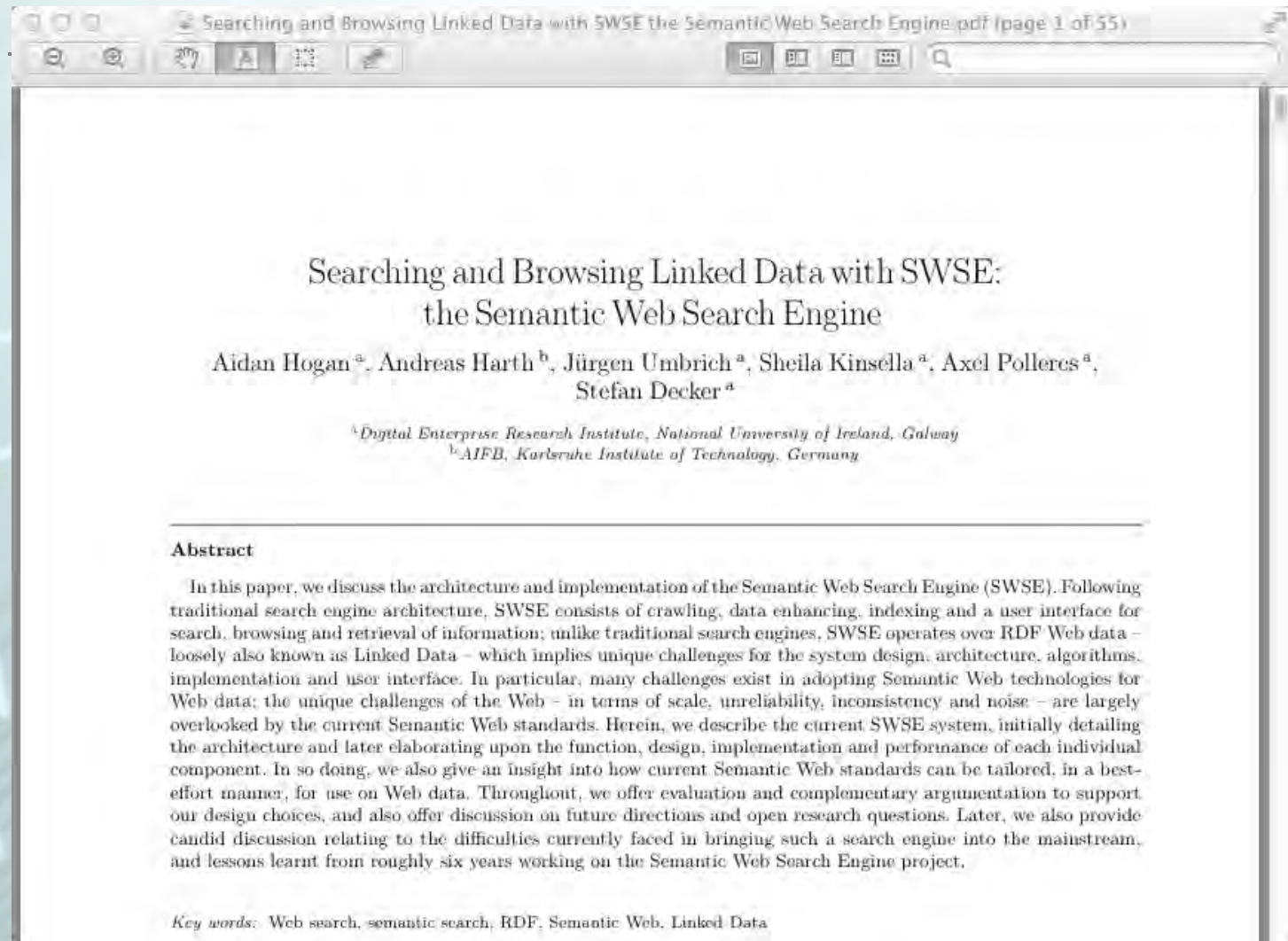
JWS BLOG

[JWS special issue: reasoning with...](#)

[Journal of Web Semantics, volume 11...](#)

[JWS special issue on scalability v10...](#)

So he could read the paper.



The paper also had...



... (low resolution) diagrams, and ...



Fig. 1. Results view for keyword query Bill Clinton

- trieves a large set of RDF data from the Web;
- the consolidation component tries to find synonymous (i.e., equivalent) identifiers in the data, and canonicalises the data according to the equivalences found;
 - the ranking component performs links-based analysis over the crawled data and derives scores indicating the importance of individual elements in the data (the ranking component also considers URI redirections encountered by the crawler when performing the links-based analysis);
 - the reasoning component materialises new data which is implied by the inherent semantics of the input data (the reasoning component also requires URI redirection information to evaluate the trust-



Fig. 2. Focus view for entity Bill Clinton

example, see [54]).

We will detail the design and operation of each of the components in the following sections, but beforehand, we present the distribution framework upon which all of our components are implemented.

2.3. Distribution Abstraction

In order to scale, we deploy each of our components over a distributed framework which we now briefly describe; Figure 4 illustrates the distributed operations possible in our framework. The framework is based on a shared nothing architecture [116] and consists of one master machine which arches-

... algorithms to read and to understand ...

Searching and Browsing Linked Data with SWSE the Semantic Web Search Engine.pdf (page 9 of 55)

- **Scale:** The crawler should employ scalable techniques, and on-disk indexing as required.
- **Quality:** The crawler should prioritise crawling URIs it considers to be “high quality”.

Thus, the design of our crawler is inspired by related work from traditional HTML crawlers. Additionally – and specific to crawling structured data – we identify the following requirement:

- **Structured Data:** The crawler should retrieve a high percentage of RDF/XML documents and avoid wasted lookups on unwanted formats: e.g., HTML documents.

Currently, we crawl for RDF/XML syntax documents – RDF/XML is still the most commonly used syntax for publishing RDF on the Web, and we plan in future to extend the crawler to support other formats such as RDFa, N-Triples and Turtle.

The following algorithm details the operation of the crawler, and will be explained in detail throughout this section.

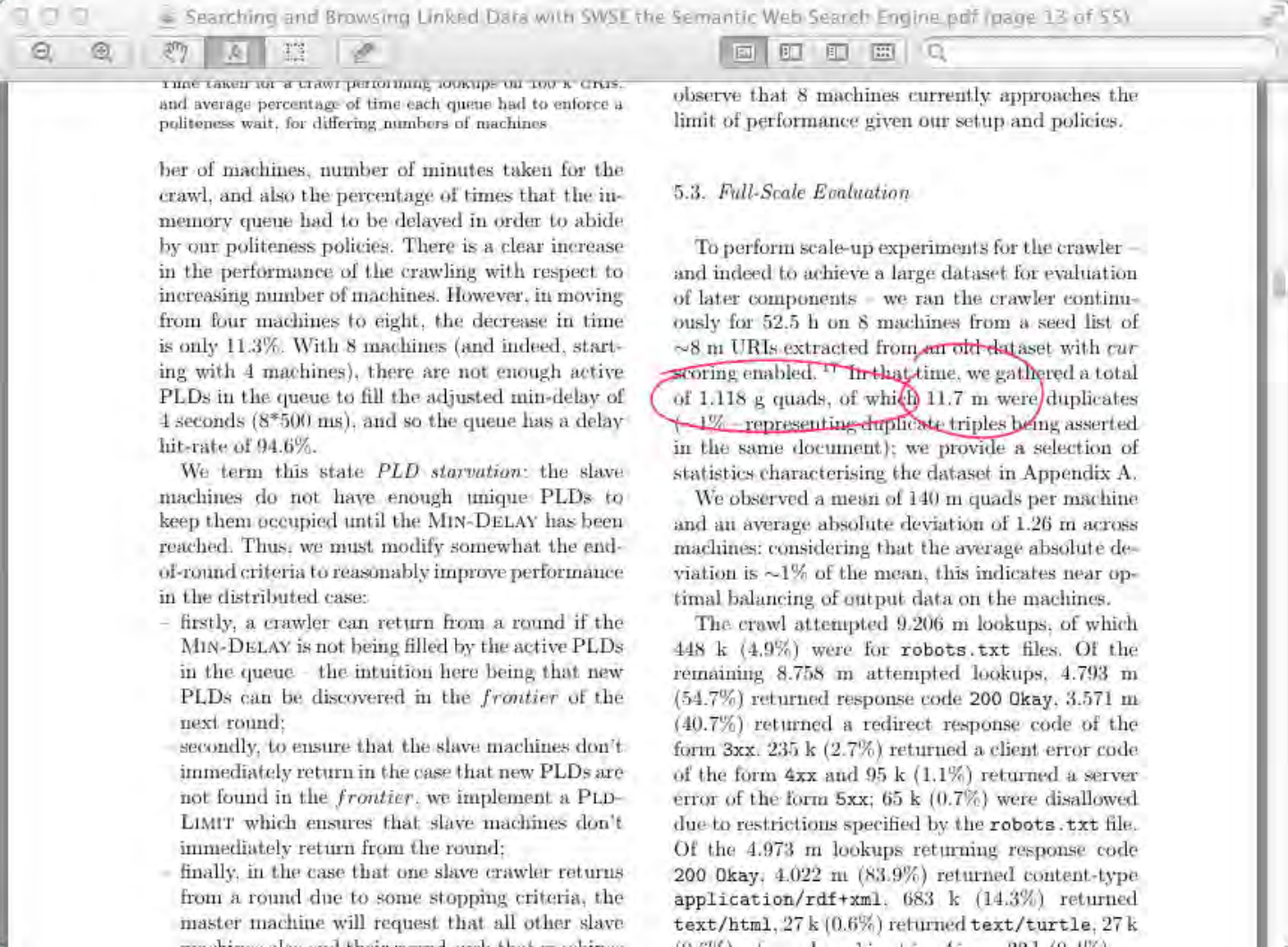
5.1. High-level Approach

Our high-level approach is to perform breath-first crawling, following precedent set by traditional Web crawlers (cf. [15] [69]): the crawl is conducted in rounds, with each round crawling a *frontier*. On a high-level, Algorithm 1 represents this round-based approach applying ROUNDS number of rounds. The frontier comprises of seed URIs for round 0 (Algorithm 1, Line 1), and thereafter with novel URIs extracted from documents crawled in the previous

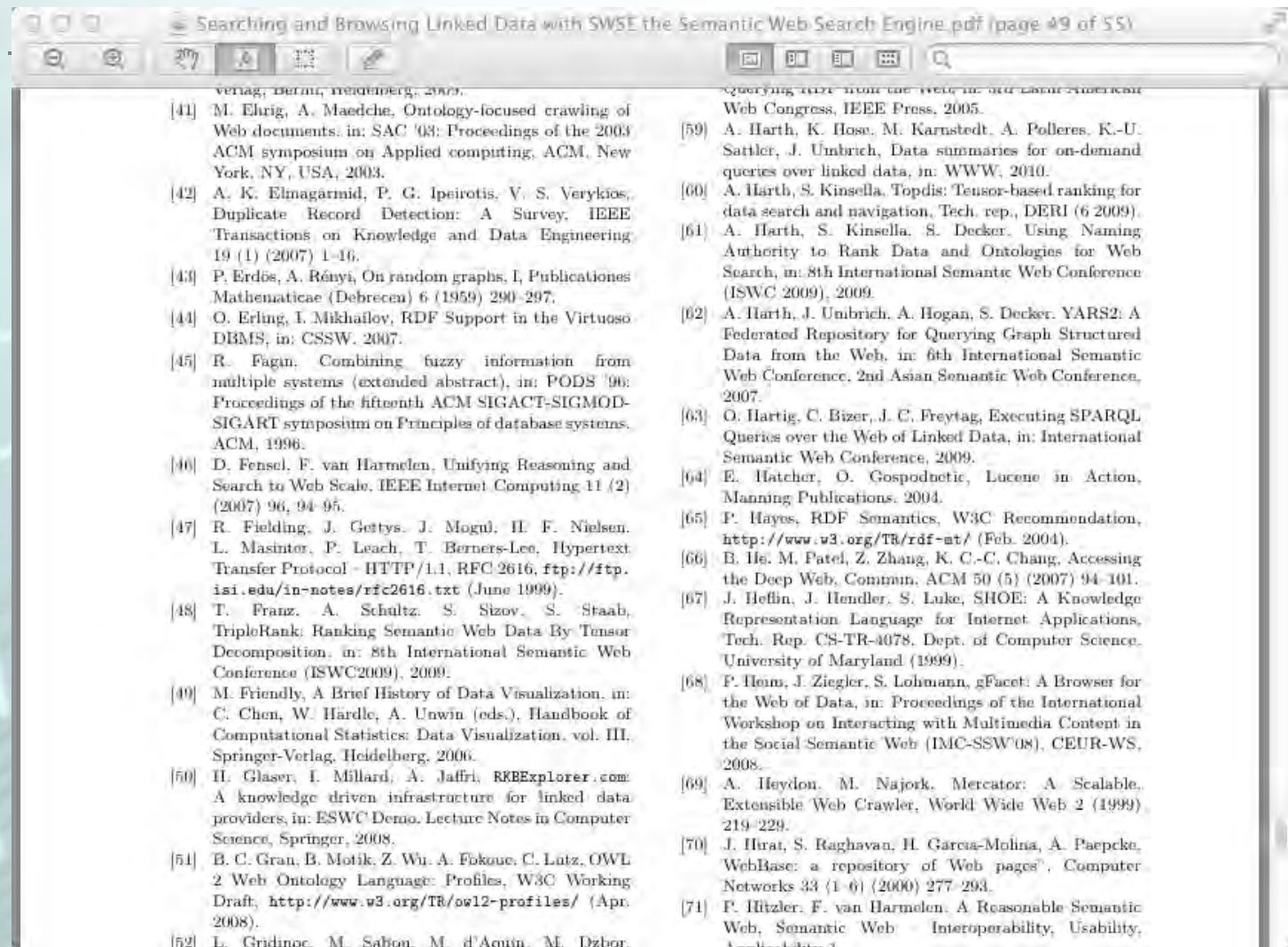
Algorithm 1 Algorithm for crawling

Require: SEEDS, ROUNDS, PLD-LIMIT, MIN-DELAY

```
1: frontier ← SEEDS
2: pld0...n ← new queue
3: stats ← new stats
4: while rounds + 1 < ROUNDS do
5:   put frontier into pld0...n
6:   while depth + 1 < PLD-LIMIT do
7:     for i = 0 to n do
8:       prioritise(pldi, stats)
9:     end for
10:    start ← current.time()
11:    for i = 0 to n do
12:      curi = calculate_eur(pldi, stats)
13:      if curi > random([0.1]) then
14:        get uri from pldi
15:        urideref = deref(uri)
16:        if urideref = uri then
17:          G = get(uri)
18:          output G
19:          UG ← URIs in G
20:          UG ← prune blacklisted from UG
21:          add unseen URIs in UG to frontier
22:          update stats wrt. UG
23:        else
24:          if urideref is unseen then
25:            add urideref to frontier
26:            update stats for urideref
27:          end if
28:        end if
29:      end if
30:    end for
```

C1=CC=C(C=C1)C(=O)N2C(=O)C(=O)N2C3=CC=CC=C3

... *lots* of references.



He liked the paper. So he wrote a blog...


Ivan's private site

http://ivan-herman.name/

2012 My Mercurial My Feedly TR LocalData Private Mailing lists Social SW Python RDFa it! Bookmarks

Ivan's private site Follow

Ivan's private site



January 24, 2012

Nice reading on Semantic Search


Filed under: [Semantic Web](#), [Work Related](#) — Ivan Herman @ 18:53 [Edit This](#)

Tags: [Linked Data](#), [OWL](#), [semantic search](#), [Semantic Web](#), [Web search engine](#)

I had a great time reading a paper on Semantic Search[1]. Although the paper is on the details of a specific Semantic Web search engine (DERI's SWSE), I was reading it as somebody not really familiar with all the intricate details of such a search engine setup and operation (i.e., I would not dare to give an opinion on whether the choice taken by this group is better or worse than the ones taken by the developers of other engines) and wanting to gain a good image of what is happening in general. And, for that purpose, this paper was really interesting and instructive. It is long (cca. 50 pages), i.e., I did not even try to understand everything at my first reading, but it did give a great overall impression of what is going on.

One of the "associations" I had, maybe somewhat surprisingly, is with another paper I read lately, namely a report on basic profiles for Linked Data[2]. In that paper Nally et al. look at what "subsets" of current Semantic Web specifications could be defined, as "profiles", for the purpose of publishing and using Linked Data. This was also a general topic at a [W3C Workshop on Linked Data Patterns](#) at the end of last year (see also the [final report](#) of the event) and it is not a secret that W3C is considering setting up a relevant Working Group in the near future. Well, the experiences of an engine like SWSE might come very handy here. For example, SWSE uses a subset of the [OWL 2 RL Profile](#) for inferencing; that may be a good input for a possible Linked Data profile (although the differences are really minor, if one looks at the appendix of the paper that lists the rule sets the engine uses). The idea of "Authoritative Reasoning" is also interesting and possibly relevant; that approach makes a lot of pragmatic sense, I wonder whether this is not something that should be, somehow, documented for a general use. And I am sure there are more: In general, analyzing the experiences of major Semantic Web search engines on handling Linked Data might provide a great set of input for such pragmatic work.

Flickr Photos



[More Photos](#)

Private

- foaf page in RDF
- foaf page in XHTML
- home
- more about me
- my latest favourite photos
- my linkedin
- my photo album
- my twitter
- PGP key

RSS

- full site atom
- full site rss 2.0

...with exact references.

The screenshot shows a web browser window titled "Ivan's private site" with the URL <http://ivan-herman.name/>. The browser's address bar and tabs are visible. The page content includes a main text area, a sidebar with navigation links, and a list of references at the bottom.

One of the "associations" I had, maybe somewhat surprisingly, is with another paper I read lately, namely a report on basic profiles for Linked Data [2]. In that paper Nally et al. look at what "subsets" of current Semantic Web specifications could be defined, as "profiles", for the purpose of publishing and using Linked Data. This was also a general topic at a [W3C Workshop on Linked Data Patterns](#) at the end of last year (see also the [final report](#) of the event) and it is not a secret that W3C is considering setting up a relevant Working Group in the near future. Well, the experiences of an engine like SWSE might come very handy here. For example, SWSE uses a subset of the [OWL 2 RL Profile](#) for inferencing; that may be a good input for a possible Linked Data profile (although the differences are really minor, if one looks at the appendix of the paper that lists the rule sets the engine uses). The idea of "Authoritative Reasoning" is also interesting and possibly relevant; that approach makes a lot of pragmatic sense, I wonder whether this is not something that should be, somehow, documented for a general use. And I am sure there are more: In general, analyzing the experiences of major Semantic Web search engines on handling Linked Data might provide a great set of input for such pragmatic work.

I was also wondering about a very different issue. A great deal of work had to be done in SWSE on the proper handling of `owl:sameAs`. On the other hand, one of the recurring discussions on various mailing list and elsewhere is on whether the usage of this property is semantically o.k. or not (see, e.g., [3]). A possible alternative would be to define (beyond `owl:sameAs`) a set of properties borrowed from the [SKOS Recommendation](#), like `closetmatch`, `exactmatch`, `broadmatch`, etc. It is almost trivial to generalize these SKOS properties for the general case but, reading this paper, I was wondering: what effect would such predicates have on search? Would it make it more complicated or, in fact, would such predicates make the life of search engines easier by providing "hints" that could be used for the user interface? Or both? Or is it already too late, because the ubiquitous usage of `owl:sameAs` is already so prevalent that it is not worth touching that stuff? I do not have a clear answer at this moment...

Thanks to the authors!

1. A. Hogan, et al., "Searching and Browsing Linked Data with SWSE: the Semantic Web Search Engine", *Journal of Web Semantics*, vol. 4, no. December, pp. 365-401, 2011.
2. M. Nally and S. Speicher, "Toward a Basic Profile for Linked Data", IBM developersWork, 2011.
3. H. Halpin, et al. "When owl:sameAs Isn't the Same: An Analysis of Identity in Linked Data", *Proceedings of the International Semantic Web Conference*, pp. 305-320, 2010

Comments (6)

Private

- foaf page in RDF
- foaf page in XHTML
- home
- more about me
- my latest favourite photos
- my linkedin
- my photo album
- my twitter
- PGP key

RSS

- full site atom
- full site rss 2.0
- sw related rss 2.0
- work related rss 2.0

Work Related

- cv, publ. list, ...
- me at W3C

Extra pages

- Some goodies to download

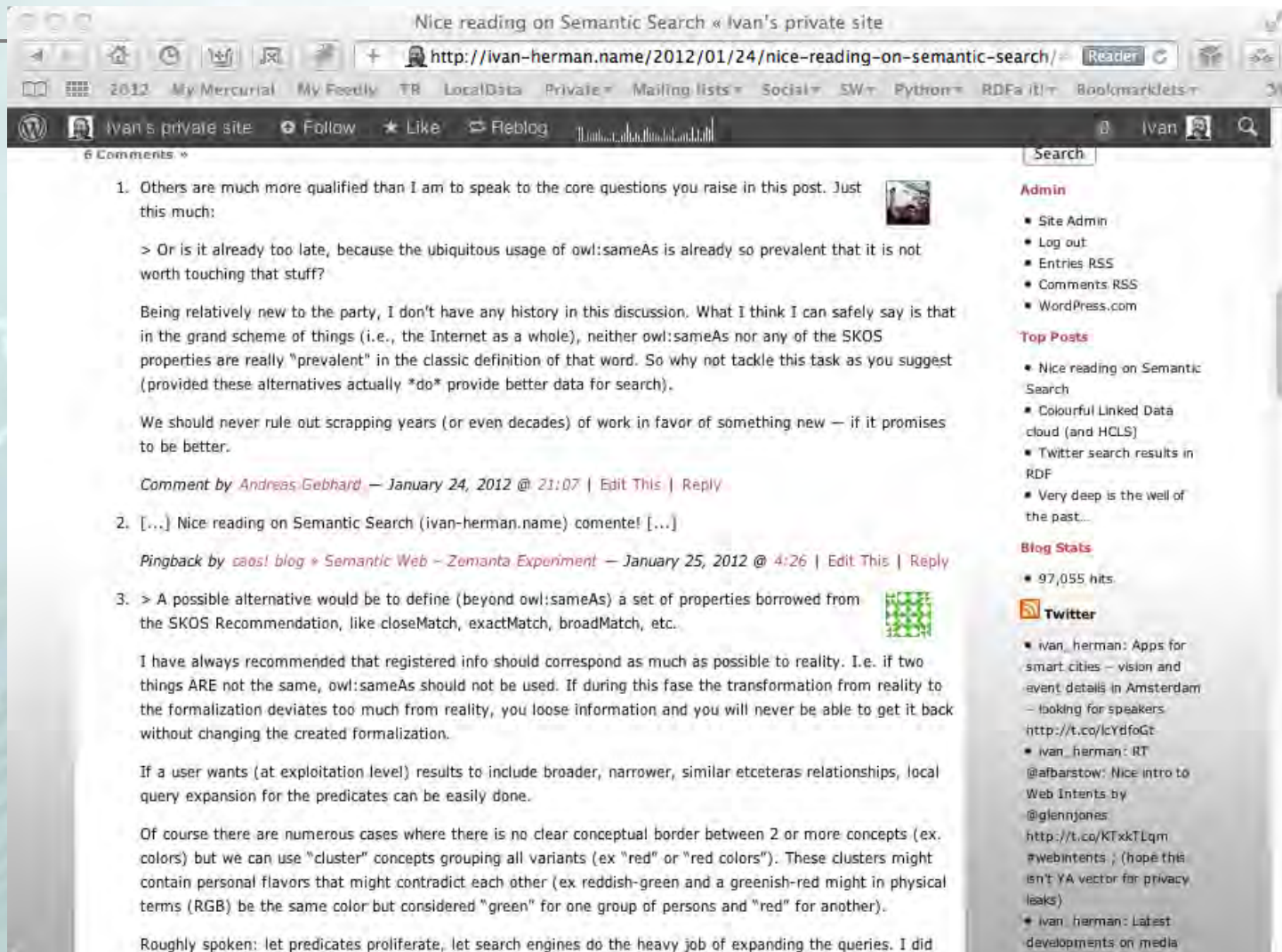
Categories

- Private (43)
- General (33)
- Hungary (14)
- Mac (1)
- Work Related (134)
- Code (15)
- Python (10)
- Semantic Web (120)
- Social aspects (5)

Archives by month

Select Month

There were a bunch of comments...



Nice reading on Semantic Search « Ivan's private site

http://ivan-herman.name/2012/01/24/nice-reading-on-semantic-search/ Reader

2012 My Mercurial My Feedly TR LocalData Private Mailing lists Social SW Python RDFa i18n Bookmarks

Ivan's private site Follow Like Reblog Ivan

6 Comments »

1. Others are much more qualified than I am to speak to the core questions you raise in this post. Just this much:

> Or is it already too late, because the ubiquitous usage of owl:sameAs is already so prevalent that it is not worth touching that stuff?

Being relatively new to the party, I don't have any history in this discussion. What I think I can safely say is that in the grand scheme of things (i.e., the Internet as a whole), neither owl:sameAs nor any of the SKOS properties are really "prevalent" in the classic definition of that word. So why not tackle this task as you suggest (provided these alternatives actually *do* provide better data for search).

We should never rule out scrapping years (or even decades) of work in favor of something new — if it promises to be better.

Comment by [Andreas Gebhard](#) — January 24, 2012 @ 21:07 | [Edit This](#) | [Reply](#)

2. [...] Nice reading on Semantic Search (ivan-herman.name) comentel [...]

Pingback by [caos! blog](#) » [Semantic Web - Zemanta Experiment](#) — January 25, 2012 @ 4:26 | [Edit This](#) | [Reply](#)

3. > A possible alternative would be to define (beyond owl:sameAs) a set of properties borrowed from the SKOS Recommendation, like closeMatch, exactMatch, broadMatch, etc.

I have always recommended that registered info should correspond as much as possible to reality. I.e. if two things ARE not the same, owl:sameAs should not be used. If during this fase the transformation from reality to the formalization deviates too much from reality, you loose information and you will never be able to get it back without changing the created formalization.

If a user wants (at exploitation level) results to include broader, narrower, similar etceteras relationships, local query expansion for the predicates can be easily done.

Of course there are numerous cases where there is no clear conceptual border between 2 or more concepts (ex. colors) but we can use "cluster" concepts grouping all variants (ex "red" or "red colors"). These clusters might contain personal flavors that might contradict each other (ex reddish-green and a greenish-red might in physical terms (RGB) be the same color but considered "green" for one group of persons and "red" for another).

Roughly spoken: let predicates proliferate, let search engines do the heavy job of expanding the queries. I did

Admin

- Site Admin
- Log out
- Entries RSS
- Comments RSS
- WordPress.com

Top Posts

- Nice reading on Semantic Search
- Colourful Linked Data cloud (and HCLS)
- Twitter search results in RDF
- Very deep is the well of the past...

Blog Stats

- 97,055 hits

Twitter

- ivan_herman: Apps for smart cities – vision and event details in Amsterdam – looking for speakers <http://t.co/lCYdfoGt>
- ivan_herman: RT @albarstow: Nice intro to Web Intents by @glennjones <http://t.co/KTxkTLqm> #webintents ; (hope this isn't YA vector for privacy leaks)
- ivan_herman: Latest developments on media

...including the author's, with his answer...



Nice reading on Semantic Search « Ivan's private site

http://ivan-herman.name/2012/01/24/nice-reading-on-semantic-search/ Reader

2012 My Mercurial My Feedly TR LocalData Private Mailing lists Social SW Python RDFa jilt Bookmarklets

Ivan's private site Follow Like Reblog Ivan

Recommendation, like closeMatch, exactMatch, broadMatch, etc.

I'm not entirely convinced that this is **needed** (yet).

Our current sketchy idea on robust use of owl:sameAs for Web data is to have different views for each individual entity, where for owl:sameAs, we would allow "transitive claims", but forego "symmetric claims" and only merge entities where reciprocal, authoritative links exist. This will imply different views for a given entity, a bit like considering owl:sameAs as a transitive "Imports" property. So if we find the triple "A sameAs B" in a location authoritative for A, the entity-view for A will include data for B, but not vice-versa. If we also find "B sameAs C" in a document authoritative for B, the view for A and the view for B will include the data for C also. If we also find the triple "B sameAs A", we can merge A and B such that they become one view. The core principle here is that a publisher has complete control over its local entities: they can opt in or opt out of what data its entities are viewed/merged with by dropping the pertinent owl:sameAs relation(s) in the local document. It would not be difficult to then support selecting and traversing these different views for a given entity in a UI.

Importantly, this doesn't require a new property or vocabulary for publishers, but rather a refined/selective/more robust interpretation of owl:sameAs by consumers. In general, when reasoning over Web data, it's more a case of disentangling who says what and where, as opposed to interpreting everything you see as formal truth. Obviously the burden is not (only) on publishers to get it right, but also on consumers to not implode when they get it wrong.

> Thanks to the authors!

And thanks for the encouraging feedback and interesting discussion!

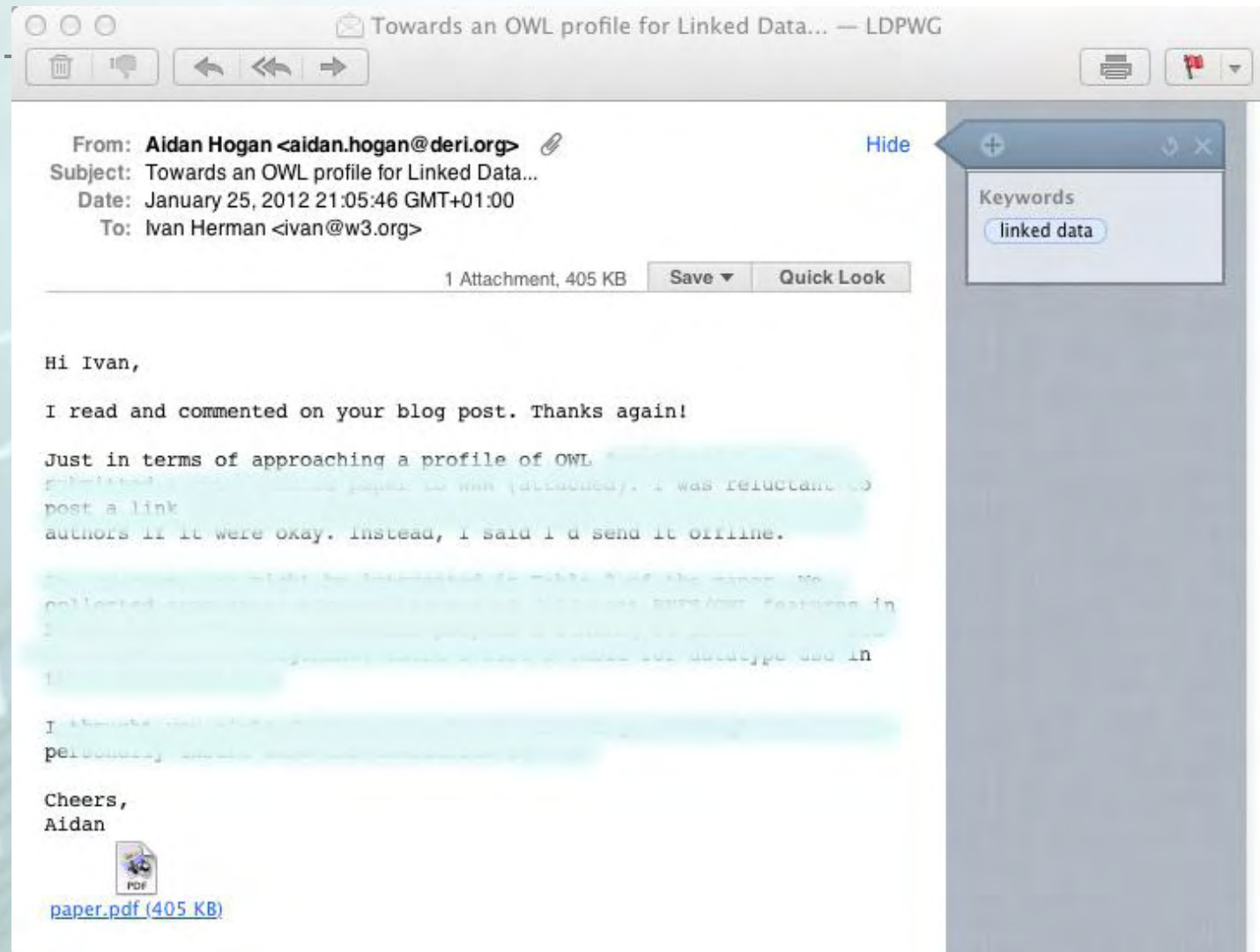
Comment by Aldan Hogan — January 25, 2012 @ 21:58 | Edit This | Reply

Wow! This is a blog post by itself, not just a comment:-) But thanks!

I think the only area where either we are in disagreement or I am not sure I understand what you say is your last section, referring to the handling of owl:sameAs. Isn't it correct that if you forego symmetric claims then, formally, you do not follow the precise owl:sameAs semantics? Of course it is perfectly for you to do that but, at least for my taste, this is not entirely ok. In other words, it is exactly for the possibility of expressing non-symmetric identifications that new predicates might be needed, to allow you to do that while staying within the definitions...

Of course, as I say, the issue might be that this train is already gone, and there are too many owl:sameAs out there.

...and also private communications



This was an example of what scholarly communication is *really* all about

- experts **find one another's result**
- they engage into private or public conversations, discussions
 - they may lead to
 - new results
 - new, possibly common actions
- **they get to know and possibly influence one another's view**
 - Etc.

Communication was important from the beginning



- Denis de Sallo and Journal de Scavans
- Henry Oldenburg and Philosophical Transactions of the Royal Society

What is different today?




Today

- cheap storage
 - Web advantages and hyperlinks
 - machine readable articles (data mining)
- scholarly journals, on the other hand, are frozen as PDF files (yes, with some links sometimes...)

Changes

- Twitter, Google+, Facebook...as a communication tools should be assessed as an impact
 - Higher pace of information exchange
 - could we justify long publication delay?

The background of the slide features faint, light blue chemical structures, including what appears to be a nucleotide base and a sugar molecule, overlaid on a light green and white gradient.

***Will scholarly communication move
away from its paper centric model, and
join the information age!***

Article of the future

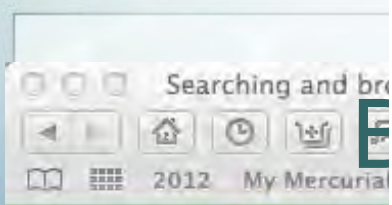
- not only text but:
 - datasets, multimedia (video, audio, images) in standardized format
 - interactivity with readers
 - social network included
 - full metadata set (date of application, date of acceptance, date of publishing, full metadata set for references list)
 - related work, previous versions of the article
 - hyperlinks usage
 - re-usable

Publishers of the future

- more flexible subscription models, copyright policies
 - new business models
 - from text to other formats
 - research data!!
 - rich content
 - more long term archiving and preservation activities
- more added value – education, promotion of the new trends

Library of the future

- from archiving and preservation to **access** and **delivery**
 - from collections to user-centric **services**
- **not to compete with, but use “big players” products** – Google, Amazon, etc., and learn from them



Stefan Decker^a

^a Digital Enterprise Research

^b AIFB, Karlsruhe Institute of

Available online 22 June 201

<http://dx.doi.org/10.1016/j.w>

[Permissions & Reprints](#)

[View full text](#)

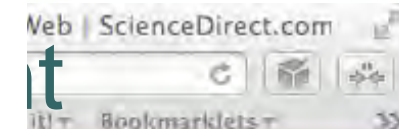


Abstract

In this paper, we discuss Following traditional sea user interface for search operates over RDF Web system design, architect exist in adopting Semar scale, unreliability, inc standards. Herein, we elaborating upon the fur doing, we also give an manner, for use on Web our design choices, and



WHY SHOULD YOU PAY TO READ THEM ?



ScienceDirect

related reference work artic
g. encyclopedias

Web Searching

Encyclopedia of Language & Linguistic

Internet

Encyclopedia of Ecology

Library Applications

Encyclopedia of Information Systems

More related reference work articles

Science should be public good

- taxpayers
 - created in non-for-profit institutions
- authors and referees don't expect financial rewards



Open access

- the best way to overcome scholarly publishing crisis
 - it's logical
 - it's natural
 - it's promising
- commercial publishers are upset – it should be good

What is open access?

■ „By Open Access, we mean the free, immediate, availability on the public Internet of those **works which scholars give to the world without expectation of payment** – permitting any user to read, download, copy, distribute, print, search or link to the full text of these articles, crawl them for indexing, pass them as data to software or use them for any other lawful purpose. ”

WHY?



Benefits

- opening the knowledge base to all - more researchers can build on it and there is less duplication of effort
- researchers can reach a greater audience and find that their work is more widely read and cited
 - institutions gain an enhanced reputation as their research becomes more visible
- funding agencies see a greater return on their investment, and publishers find that the impact of their journals increases.

Greater visibility and accessibility, and impact

- according Harnad and Brody study open access doubles downloads and increases citations by an average of around 50% (from around 40% for biology to 250% for physics)
- 20% research in open access – a lot of impact (money invested in R&D) has been lost every year

Rapid and more efficient progress of scholarly research

- an example of arXiv with 768,539 papers
 - 1.5 million of connections per day
- time taken for articles to be cited is shorter every year

The background of the slide features faint, overlapping chemical structures. On the left, there is a complex molecule with a pyrimidine-like ring system, possibly a nucleotide or a related compound. To the right, there is a long, zigzag chain representing a polymer, likely a protein or a synthetic polymer. The structures are rendered in a light, sketchy style, blending into the light blue background.

HOW?

Two roads

- network of institutional repositories (archives) + metadata exchange = green road
 - journals in open access = golden road

Institutional repositories...

- “digital collections that preserve and provide access the the intellectual output of an institution.”
 - selfarchiving
 - mandate

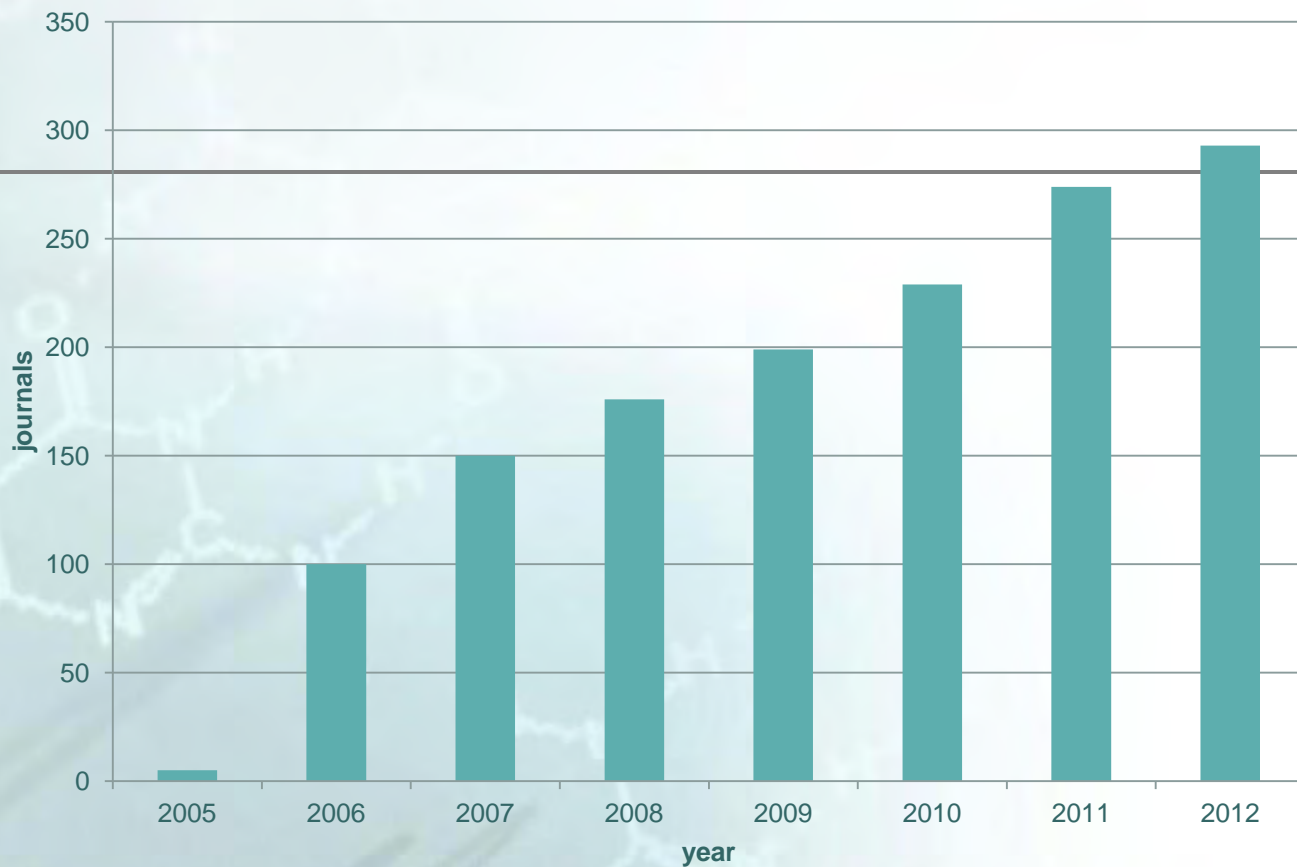
Raym Crow. *The case for institutional repositories: a SPARC position paper*. 2002

Open access journals

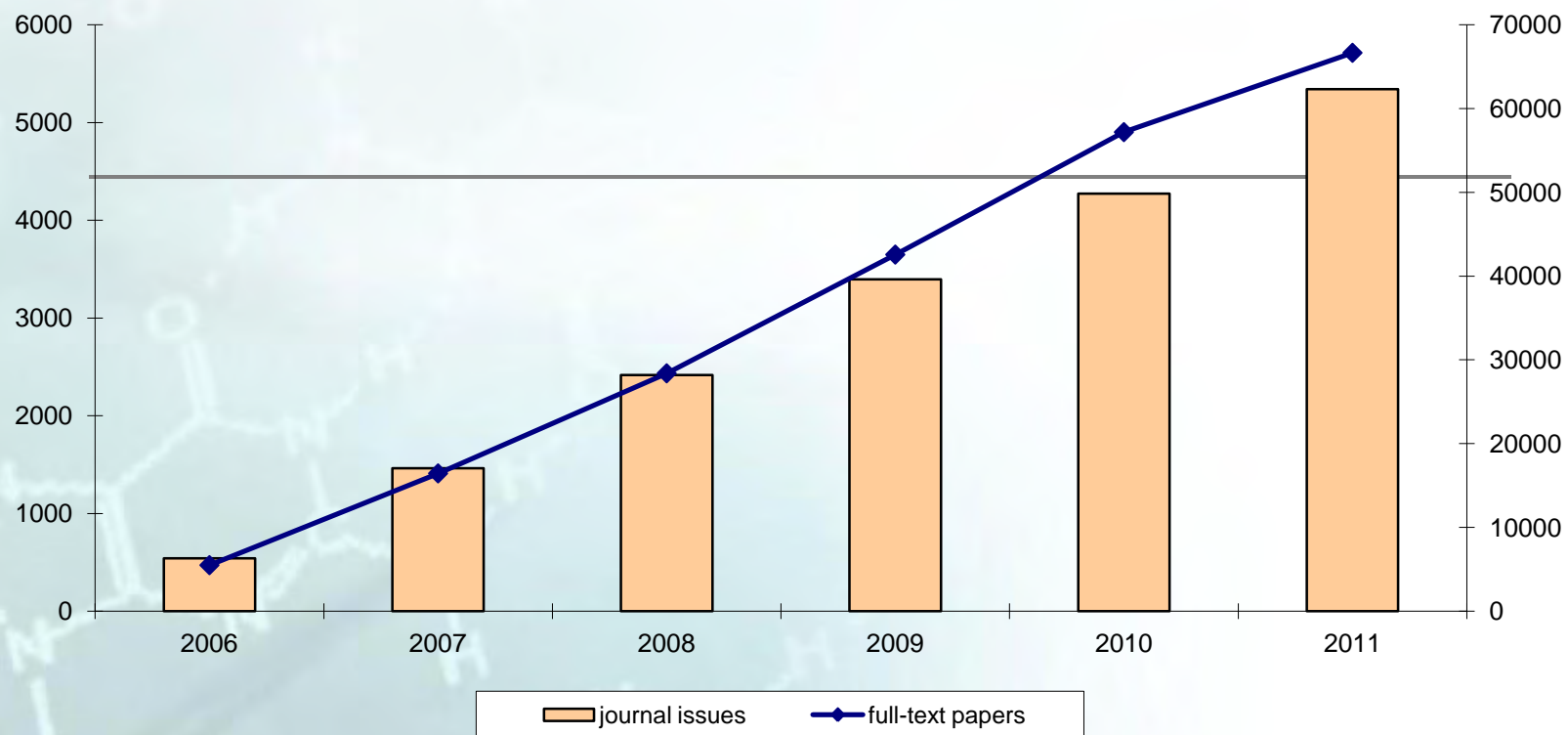


HRČAK – initial goals

- simple way to make online version of the (printed) journal
 - single access point for all Croatian open access journals (scholarly, professional and popular science)
 - metadata and full-text articles repository
- data sharing – with international repositories, databases, archives



number of HRČAK journals by years



**Number of journal issues an full-text papers
by year**

Journal	Inclusion date	Full-text visits	Total visits
Bogoslovska smotra	10.1.2009	373465	715568
Društvena istraživanja	11.4.2006	300454	682859
Ekonomski pregled	6.2.2007	391306	582121
Collegium	10.7.2006	247850	388749
Antropologicum			
Građevinar	11.1.2007	248931	361871
Acta stomatologica	7.2.2006	209667	347778
Croatica			
Politička misao	20.2.2008	183359	335451
Narodna umjetnost:	28.4.2006	182708	328320
hrvatski časopis za			
etnologiju i folkloristiku			
MEDICUS	11.1.2008	208905	326598
ŽIVOT I ŠKOLA: časopis za	14.3.2008	248036	320706
teoriju i praksu odgoja i			
obrazovanja			
Acta Clinica Croatica	10.8.2007	118985	301218

2010-2012

Data harvesting

- OAI-PMH protocol
 - [OAISTER](#)
 - [BASE](#)
- [Google Scholar](#)
- ...

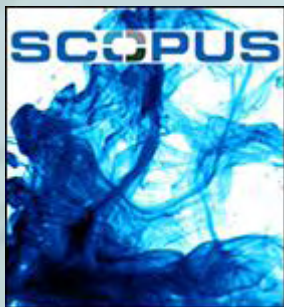
Harvesters visited HRČAK on average:

- 2508 times per month in 2011
- 3543 times per month during the first four months in 2012



Cooperation with commercial database publishers

- Elsevier/Scopus – 105 Croatian journals
- ThomsonReuters/WoS – 65 Croatian journals
 - Google Scholar – **all HRČAK journals**
 - citation impact



Web 2.0 applications

Kemija u industriji, Vol.59 No.4 Travanj 2010.

Stručni rad



From Our Libraries:

Danko Škare (ed.)

Metrics of Scientific Journalism - Truths, Myths and Misconceptions

J. Stojanovski; University of Zadar, Zadar, Hrvatska; Ruđer Bošković Institute, Zagreb, Hrvatska

[Puni tekst \(Hrvatski\) Str. 179 - 186 \(pdf, 1.09 MB\) downloads: 482](#)

Sažetak

The evolution of the digital age has led to the development of so-called abstracting/indexing printed publications in online databases. A short survey of the four most popular databases in Croatia has been done, as Current Contents, Web of Science, Scopus and Google Scholar, with short history, coverage (included Croatian titles), features and possibilities of citation analysis. We discuss their strengths and weaknesses, and the importance of the efficient usage and appropriate interpretation of the data. The journal impact factor and other bibliometric measures are frequently misused to estimate the influence of individual papers and authors. Explaining potential benefits and limitations of bibliometric indicators, this article may help policymakers, librarians, administrators, and individual researchers to use the data provided by different databases more carefully and wisely in the evaluation of research.

Ključne riječi

metrics: scientific journalism; citation analysis; online databases; impact factor

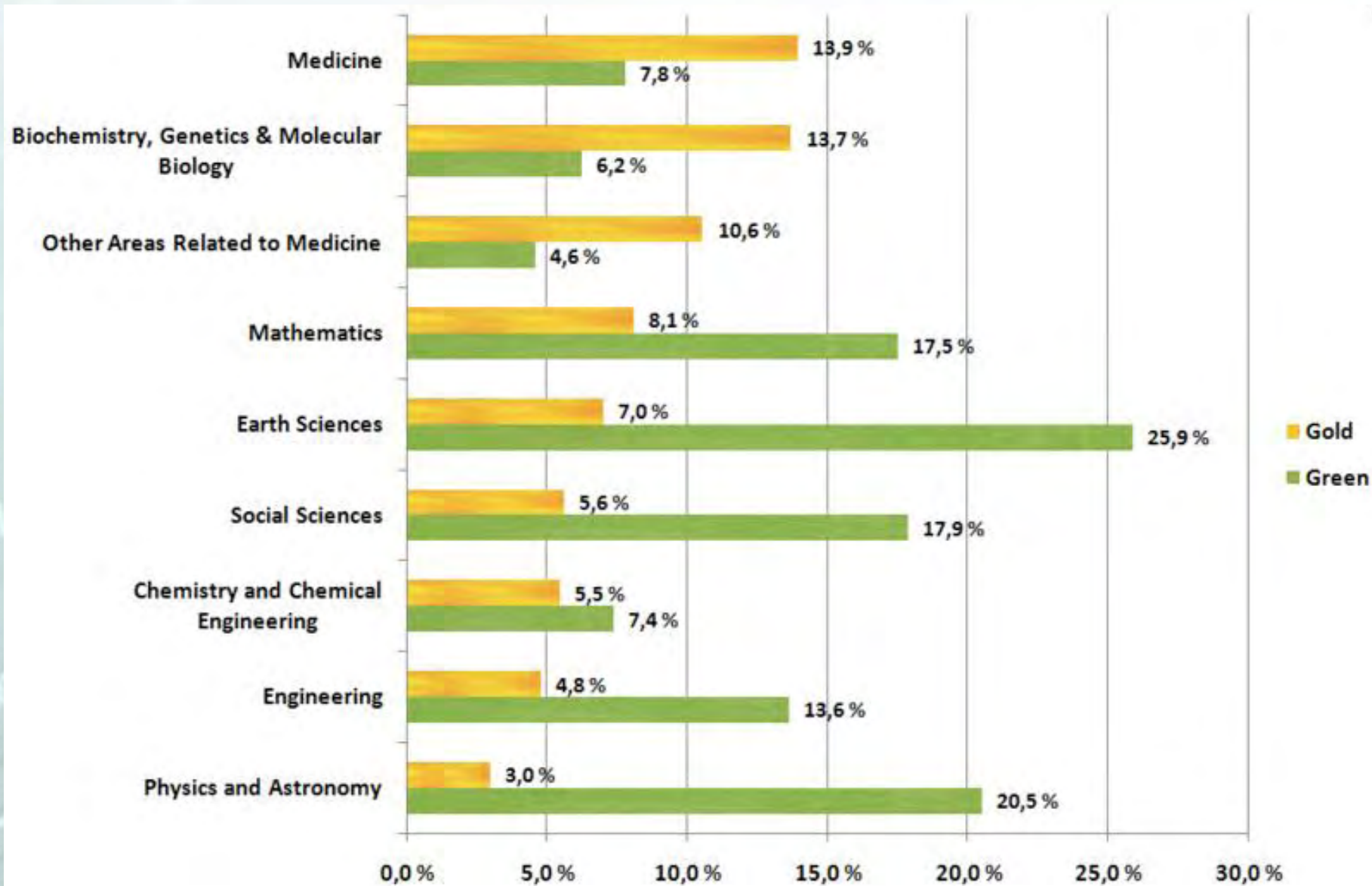
[\[Hrvatski\]](#)

Posjeta: 1031 (od 01.01.2007.)

Fifth Belgrade International Open Access
Conference 2012, Belgrade, May 18th 2012

How many journals are in OA

- **8.5%** free available at publishers web sites
= **golden OA 8.5%**
- additional papers in the institutional repositories and authors' and institutions' web sites = **green OA 11.9%**
= **20.4 %** annual scholarly output is freely available



The background of the slide features faint, light blue chemical structures, including what appears to be a nucleotide base and a peptide chain, overlaid on a light green and white gradient. A thin horizontal line is positioned above the text.

WHERE? EVERYWHERE.

Where to publish or what makes great journals ?

- it is not about technology, big investments, or **great promotion**
 - journals are based on the communities they serve
- living organism that rely on the editors, authors and reviewers that make up that community
 - no value without the active support of high level scientists
 - quality attracts quality

How do authors choose a journal?

- subject coverage of their research paper and its quality and approach
- selection of the most appropriate journals in terms of subject coverage and readers
- matching the general quality of their paper (best, good, ok) to a class of journals (top, average, low)
 - from that class they select a specific journal based upon experience
 - recommendation from professor

How do author choose a journal?

Journal metrics (IF etc.)

Reputation

Editorial standard

Publication speed

Access to audience

International coverage

Self evaluation

A&I coverage

Society link

Peer Review?


Metrics?

- **existing indicators are often misused**
- **„wrapping paper“ is more important than a content**
 - impact factor
 - SJR
 - SNIP
 - eigenfactor
 - h-index
 - article influence
 - number of citations
 - number of papers
 - ...

Possible metrics for an article

- total cites (Google Scholar, Scopus, WoS) – deduplicated
 - # visits
 - # downloads
 - # comments
 - # bookmarks, likes...
 - expert assessments...
- # discussions (on well known blogs)
- # appearance in news, blogs, etc.

 **Download:** [PDF](#) | [Citation](#) | [XML](#)

 **EzReprint** New & improved!

OPEN  ACCESS

Id: The Identification of the Two Missing Children Using DNA Analysis

ics

Related Content

Comments: 3

Metrics

Total Article Views: [48,368](#)

Citations

[CrossRef \(14\)](#)

[PubMed Central \(6\)](#)

[Scopus \(15\)](#)

[Web of Science® \(15\)](#)

Social Networks

[CiteULike \(2\)](#)

[Facebook \(282\)](#)

[Mendeley \(19\)](#)

[More](#)

Related Content

Related Subject Categories

[Chemistry](#), [Genetics and Genomics](#), [Molecular Biology](#)

Related Articles on the Web

[Google Scholar](#)

[PubMed](#)

[More](#)

Share this Article



 [Email this article](#)

	HTML Page Views	PDF Downloads	XML Downloads	Totals
PLOS	43,318	2,252	74	45,644
PMC	2,081	643	112	2,724

CLE

OPEN  ACCESS

Id: The Identification of the Two Missing Children Using DNA Analysis


etrics

Related Content

Comments: 3

[Comment on this article](#)

2 [As far as I understand, this conclusion in Discussion" is not based on DNA-tests. Probably, this is only his opinion.](#)

Posted by  [BorisRomanov](#) on 02 Mar 2011 at 11:39 GMT

Use on 02 Mar 2011 at 11:39 GMT

2 [Additional support to GMI by the FWF Austrian science Fund](#)

Posted by  [coblem](#) on 11 Mar 2009 at 22:28 GMT

Use on 11 Mar 2009 at 22:28 GMT

2 [Additional support to GMI by the FWF Austrian science Fund](#)

Posted by  [coblem](#) on 11 Mar 2009 at 22:28 GMT

Use on 11 Mar 2009 at 22:28 GMT

Parts of presentations (with permissions)

- The story from the beginning is from: Ivan Herman: Report on the **"Future of Research Communications"** Workshop, Dagstuhl, August 15-18, 2012
 - slides 52-54: Anne Kitson, Executive Vice President, Health and Medical Sciences, Elsevier: Editor Seminar in Journal Publishing - Attaining Excellence in Scholarly Communication

<http://www.javatuning.com/my-first-question-at-stackoverflow-com/>



THANK YOU! 😊