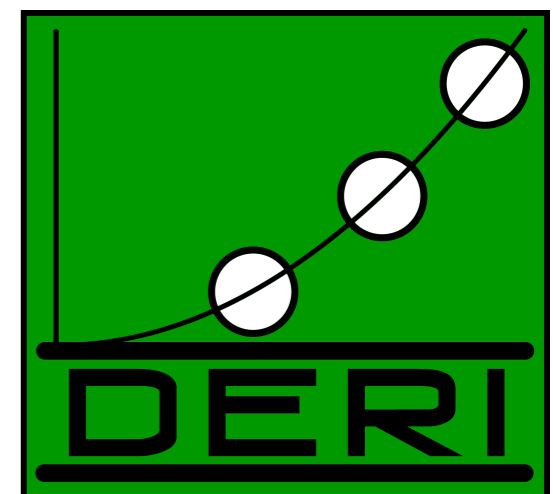


Sindice.com

Richard Cyganiak



Web of Data

- Linked Data and microformats
- 50M+ documents
- Social graph, geolocating, tags, open data
- Special-interest community data

**To use this data,
you have to find it first**

Where to find...

- ...documents that talk about X?
- ...good identifiers for Y?
- ...different URI aliases for Z?



Sindice - Data Web Services

Over 10 billion pieces of reusable information can already be found across 100 million web pages which embed RDF and Microformats. Start consuming this data today with Sindice Data Web services.

[LEARN MORE →](#)

Keyword

URI

Property

Advanced

Search the Semantic Web

SEARCH

Examples: [tim](#), [michele](#), [deri](#)

Now searching index V1 (0 documents and counting). Also try [index V0](#) (26.9 M)

Latest News

Meet us at:



XTech: May 6-9, Dublin, Ireland

SemTech: May 18-22, San Jose, California

Linked Data Planet: June 17-18, New York City

SINDICE BLOG

 [Feed](#)

[An Exciting Hard Hat area](#)

MAY 30, 2008

As you might have noticed by the look of the site we're now in a very exciting transition phase. Here is a short summary of the ma...

[\(More →\)](#)

[Sindice in use](#)

FEB 20, 2008

Sindice is really meant to be used by your project, and for us it couldn't be any nicer than seeing this happening more and more :-)

[\(More →\)](#)

[Sindice @ 20+ Millions and Openings](#)

NOV 22, 2007

Works on Sindice are proceeding at full speed and so is the indexing of the Semantic Web.

[\(More →\)](#)

[Sindice Search in your Web Browser](#)

NOV 07, 2007

What we do

- Crawl the Web for RDF and microformats
- Index based on keywords, URIs used, ...
- Offer search interface
- Offer APIs for app developers

What we don't do

- Triple store, SPARQL
- Full content
- We tell you where it is
- You fetch & parse & process

Basic API operations

- `find documents(keyword): URI[]`
- `find documents(mentioned URI): URI[]`
- `find documents(IFP, value): URI[]`

Really Soon Now

- find documents(triple pattern): URI[]
- any boolean combination of the above
- restrict on domain, ontology, format, class
- query inferred triples (quarantined reasoning)

In the pipeline

- find entities(...)
- real-time smushing
- ...

Under the hood

- 20-box cluster
- Everything distributed, Hadoop
- 7M documents indexed (30M in old index)
- Java backend, Ruby on Rails frontend
- Lots of magic

Who's using it?

- SIOC widget
- SWAML
- Meaning Of A Tag project
- Semantic Web Client Library
- Python and Prolog libraries

Contribute!

- Send us pings
- Publish Semantic Sitemaps
- forum.sindice.com



<http://sindice.com/>