



PERPUSTAKAAN NASIONAL  
REPUBLIK INDONESIA



# Indonesia OneSearch: NLP, Text Analysis, dan Peluang Baru bagi Pustakawan

**Ismail Fahmi, PhD.**

Inisiator Indonesia OneSearch (IOS)

[Ismail.fahmi@gmail.com](mailto:Ismail.fahmi@gmail.com)

**Seminar Nasional Kepustakawanan**

Perpustakaan UGM, Yogyakarta

30 November 2016



## Ismail Fahmi, PhD.

Ismail.fahmi@gmail.com

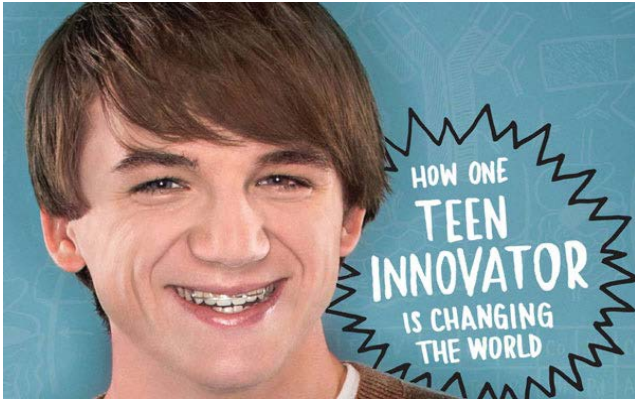
- 1992 – 2007 S1, Teknik Elektro, ITB
- 2003 – 2004 S2, Computational Linguistics, Universitas Groningen, Belanda
- 2004 – 2009 S3, Computational Linguistics, Universitas Groningen, Belanda
  
- 2000 – 2003 Inisiator IndonesiaDLN (Digital Library Network pertama di Indonesia)  
Mengembangkan Ganesha Digital Library (GDL)  
Mendirikan Knowledge Management Research Group (KMRG) ITB  
Membangun Digital Library ITB
  
- 2009 – Sekarang Engineer di **Weborama**, Perusahaan berbasis big data (Paris/Amsterdam)
- 2012 – Sekarang Co-Founder **Awesometrics**, Media Monitoring & Analytics Company
- 2014 – Sekarang Founder **PT. Media Kernels Indonesia**, a Natural Language Processing Company
- 2015 – Sekarang Konsultan **Perpustakaan Nasional**, Inisiator Indonesia OneSearch

# Agenda

---

- Indonesia OneSearch
- Knowledge Graph
- NLP
- Text Analysis and Advanced Visualization
- Peluang Baru bagi Pustakawan

# Kisah Jack Andraka dan Open Access



- Menciptakan alat deteksi kanker pankreas (Steve Jobs meninggal karena penyakit ini).
- Alat ciptaan Jack harganya sangat murah (3 cent), akurasi sangat tinggi, dan bisa dipersonalisasi untuk kondisi pasien.
- Ketika riset, Jack sering terbentur oleh payment wall dari repositori berbayar.
- Jack memanfaatkan Google untuk mencari dan mendapatkan artikel open access.
- Open Access telah memudahkan penemuan baru dan inovasi.

# OPEN ACCESS: BENEFITS




**Increased Visibility**



**Increased Citation Rates**



"Jack Andraka, a 15 year old from USA, used free articles from Google Scholar to develop a highly innovative method of detecting a rare type of pancreatic cancer."



**Drives Innovation**



**Global Impact**

"Piya Sorcar, founder of Teach Aids made materials & research about HIV available for free globally. The material is now available in 15 languages and has reached over 70 countries, including numerous educators, governments and NGOs."

Open Access brings new audiences to scholarly content, such as educators, business owners, support workers and so on by removing barriers of affordability and access.



**Public Access**



**Comply with Funder Policies**

Funders, such as the Research Councils, European Commission and the Wellcome Trust, mandate for open access publication of research outputs.

# Open Access Around The World

**OpenDOAR**

**DOAJ** DIRECTORY OF  
OPEN ACCESS  
JOURNALS

**OpenRepository**



Open Access



Scientific  
Open Access  
**SOAJ** Journals



**SPARC\***

**ELSEVIER**

**OPEN ACCESS** |  MALAYSIA

**Open access**

 **OAK**  
OPEN ACCESS KOREA

 **PLOS**

**OSJI** | OPEN  
SCIENCE  
JOURNAL

**OPEN ACCESS**

**Karger**  
**Open access**

**open india**

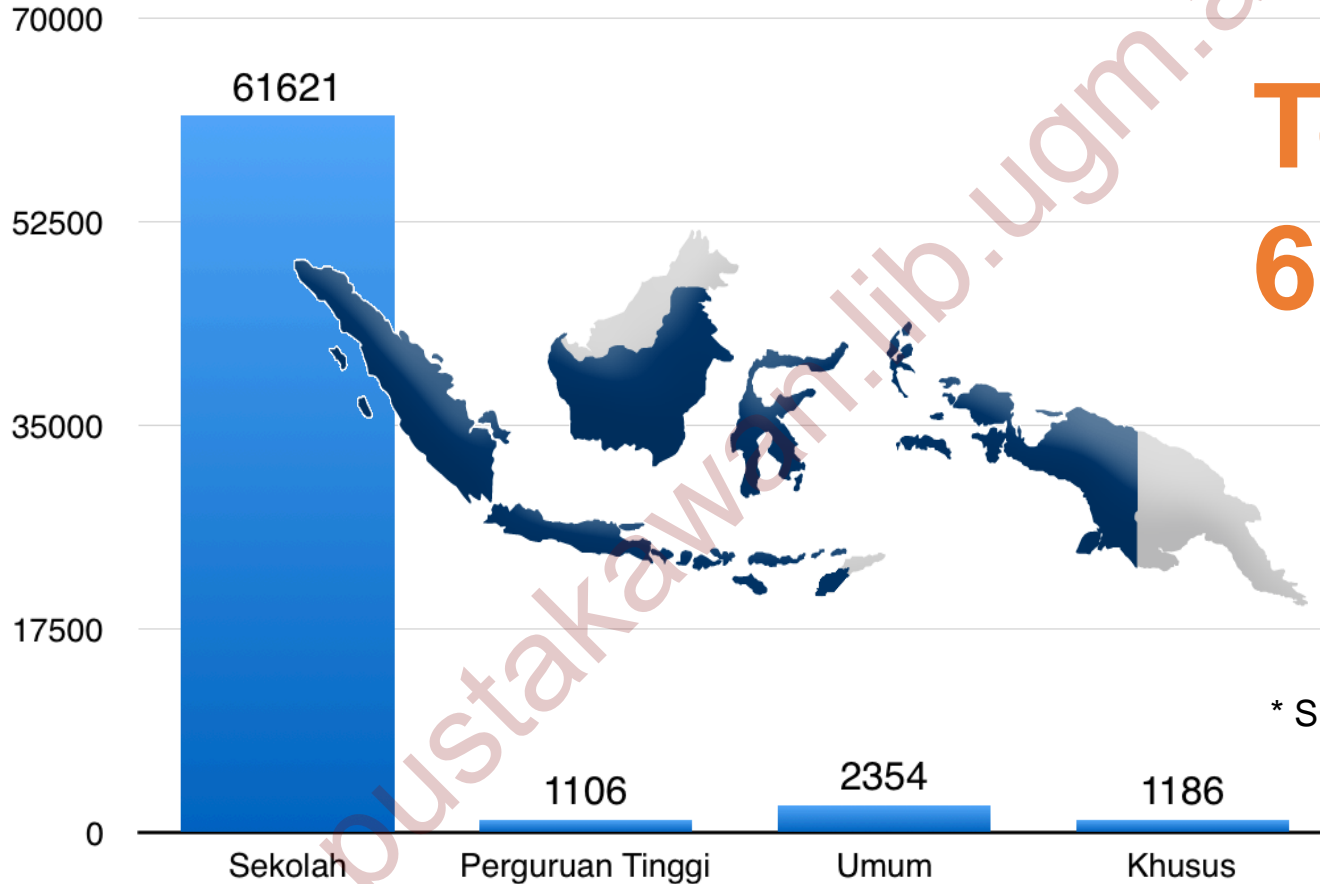
**doab** | directory of  
open access  
books

# Indonesia OneSearch

Introduction

pustakawan.lib.ugm.ac.id

# Perpustakaan di Indonesia








**Total:  
66.000+**

\* Sumber: Perpunas 2016



# Discovery System di Indonesia

	Katalog Buku	E-Journal	Digital Repository	Multi Sektor
	●			
		Discontinued		
		●		
		●		
	●	●	●	●

# OAI Discovery System



Merupakan sebuah



# Road Map IOS

## Tahap 1 (2015): OneSearch Portal

### OneSearch Portal (Bibliografi)

- Software Indonesia OneSearch
- Harvesting data bibliografi
- Protokol standard OAI-PMH
- Repository: Katalog buku, Jurnal Online, Repositori digital



## Tahap 2 (2016-2017): Text Analysis

### Text Analysis (Full Teks)

- Crawling fullteks (PDF) TA, tesis, disertasi, laporan peneliti, dan artikel jurnal dari Intitusi di Indonesia.
- Text analysis menggunakan teknologi NLP (Natural Language Processing)
- Information Extraction & Knowledge Mapping berbasis NLP
- Research Mapping antar Institusi

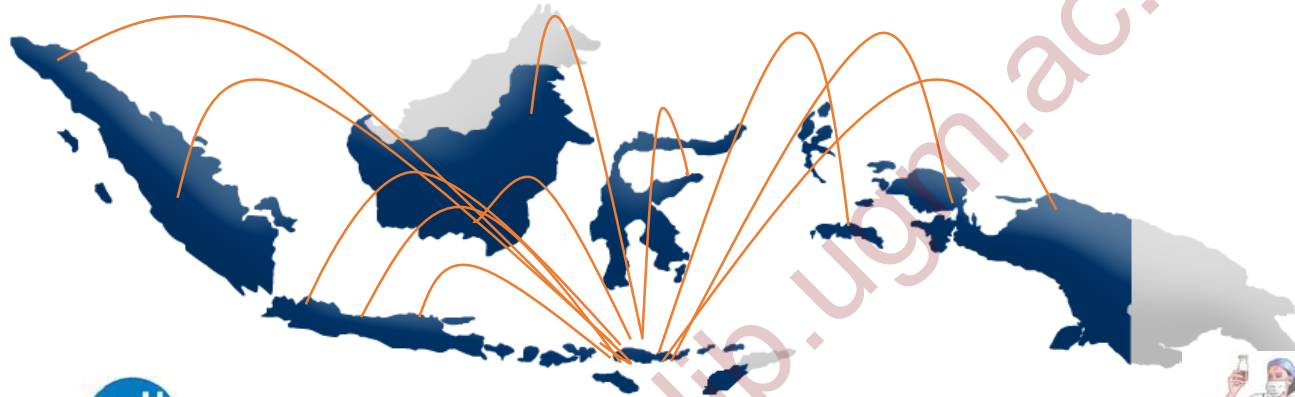


## Tahap 3 (2018-2020): Layanan Anti Plagiarism

### No Plagiarism (Services)

- Sistem dan Layanan NoPlagiarism untuk karya berbahasa Indonesia.
- Sumber: Wikipedia (Bahasa Indonesia), Online News, TA, Tesis, Disertasi, artikel jurnal, laporan penelitian (open access)
- Layanan online plagiarism checking untuk mahasiswa dan peneliti di Indonesia





Any platforms

Any collections

**iScholarly** **gdl**

**inlis** **koha** **koha**  
Integrated Library System LIBRARY SYSTEM

**Senayan** **CDS/ISIS**  
LIBRARY AUTOMATION

**greenstone digital library software**

**DSPACE** **MySIPISIS PRO**  
MARC MySQL Web-based SIPISIS

**eprints**  
repository software

**Laporan Penelitian**

**Buku, E-book**

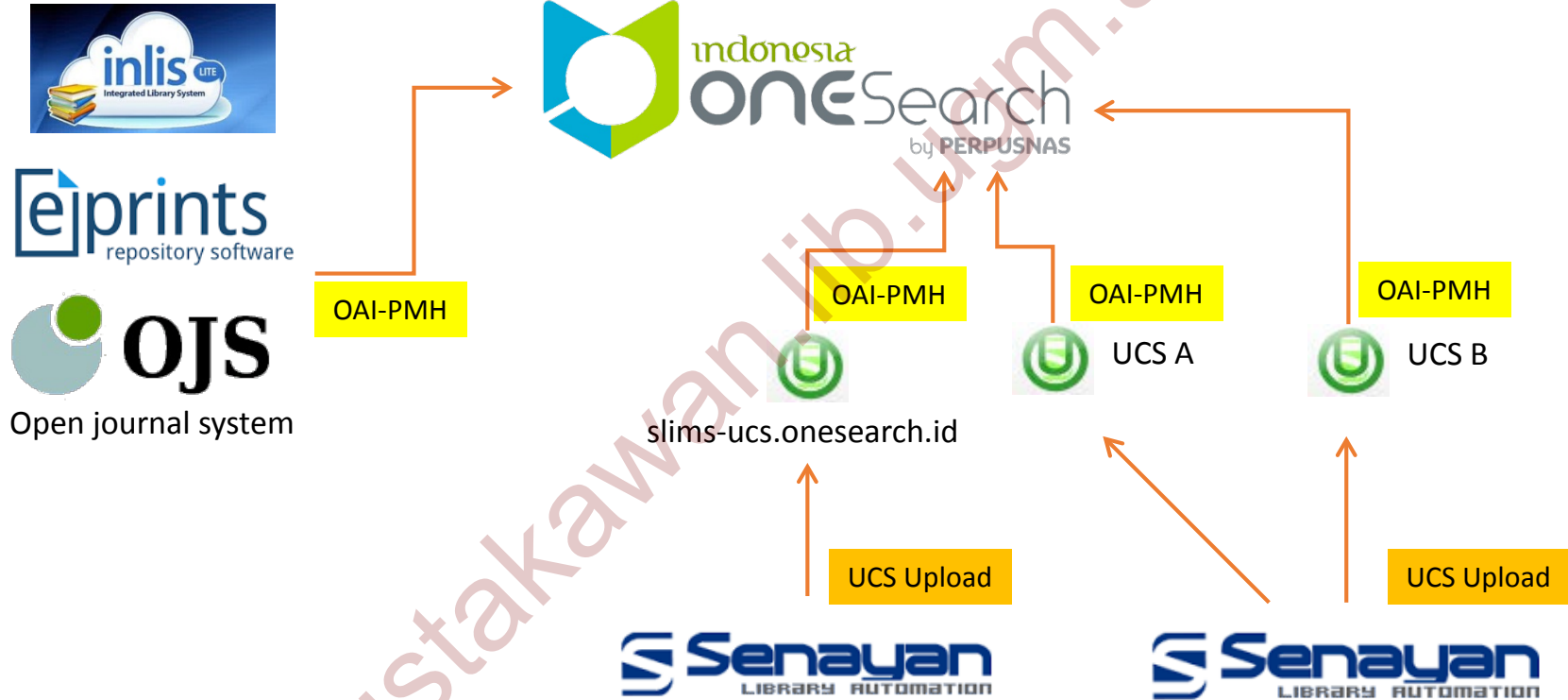
**Skripsi, TA, Tesis, Disertasi**

**Electronic Journals**

**Digitalisasi Koleksi Langka**

pustakawati@lib.ugm.ac.id

# Interoperability





The screenshot shows the homepage of the Indonesia OneSearch website. At the top, there is a navigation bar with a home icon, 'Tentang IOS', 'Organisasi Mitra', 'Masuk', and 'Bahasa Indonesia'. Below this is a blue header with the OneSearch logo and the tagline 'Satu pintu pencarian untuk semua koleksi publik dari perpustakaan, museum, arsip, dan sumber elektronik di Indonesia'. A search bar contains the text 'Cari buku dan e-resources'. To the right of the search bar are links for 'Semua Kolom', 'Cari', and 'Lanjutan'. The main content area is divided into several sections: 'Selamat Datang' with a welcome message, 'Koleksi' with statistics on the number of entries and a list of sectors, 'Mitra' with icons for 'Institusi (418)', 'Perpustakaan (454)', and 'Repositori (3171)', and 'Fitur-fitur' with icons for 'Satu Interface', 'Anti Duplikat', 'Pencarian', 'Analisa Teks', 'Jendela', and 'Integrasi'. On the right side, there are two vertical lists: 'Sektor' and 'Format', each with a list of categories and their respective counts.

**Sektor**

- Library (2,846,612)
- University (1,570,322)
- National (716,522)
- Public (501,036)
- Special (53,366)
- School (5,366)

**Format**

- Book (1,730,316)
- Other (709,441)
- Journal (577,975)
- Thesis (357,791)
- PeerReview (345,568)
- File (318,354)
- Map (50,868)
- Image (24,952)
- Sound (24,241)
- Report (13,299)
- Document (5,990)
- WorkOfArt (2,136)
- Video (759)
- Place (1)

### Top Contributors

**784,998**

Perpustakaan Nasional RI

**188,323**

BPAD Daerah Istimewa  
Yogyakarta

**151,566**

Universitas Lambung  
Mangkurat

**135,731**

Universitas Andalas

**124,731**

Badan Arsip Dan  
Perpustakaan Jawa Tengah

**111,130**

Universitas Syiah Kuala

**101,836**

Universitas Gadjah Mada

**90,575**

Universitas  
Muhammadiyah Surakarta

**85,355**

Telkom University

**74,024**

UIN Sunan Ampel  
Surabaya

**65,965**

Universitas Negeri Malang

**63,431**

UIN Syarif Hidayatullah  
Jakarta

**61,989**

Badan Perpustakaan Prov.  
Kaltim

**56,976**

Universitas Islam Negeri  
Maulana Malik Ibrahim  
Malang

**52,100**

Universitas Diponegoro

**40,151**

Universitas Sumatera Utara

**38,579**

Universitas Kristen Petra  
Surabaya

**36,910**

Institut Pertanian Bogor

**28,741**

Kantor Perpustakaan dan  
Arsip Daerah Kota  
Bandung

**27,419**

UIN Alauddin Makassar

**26,668**

Universitas Ahmad Dahlan

**26,515**

Universitas 17 Agustus  
1945 Semarang

**25,734**

Perpusda Kabupaten  
Probolinggo

**25,413**

Universitas Surabaya

**25,398**

Perpustakaan Daerah Kota  
Salatiga

**24,969**

Universitas Bosowa  
Makassar

**24,556**

Universitas Terbuka

**24,267**

Universitas Bina Nusantara

**24,262**

KPAD Kota Pekalongan

**23,368**

IAIN Surakarta

# Powerful Search Engine

Semua Kolom ▾ Cari Lanjutan

**Cari:** internet infrastructure

Koleksi Nasional

E-resources

**Persempit Pencarian**

Hapus Filter

Format: File:application/pdf ✕

Sektor ▲

Library	151
University	93
Special	58

Institusi ▲

Eprints in Library and Information Science (E-LIS)	55
Universitas Diponegoro	22
Universitas Riau	9
Universitas Ahmad Dahlan	5
Universitas Brawijaya	5
Universitas Sam Ratulangi	5

Menampilkan 1 - 20 of 166 untuk pencarian: 'internet infrastructure', lama mencari: 0.48s

NO IMAGE AVAILABLE

☰ Daftar ☰ Grid ☰ Visual

**Sortir** Relevance ▾

**Strengthening rural information infrastructure through e-Choupals**

Proceeding
PeerReviewed
application/pdf

oleh **Das, Anup Kumar**  
 Terbitan: IASLIC, 2004  
 Institusi: Eprints in Library and Information Science (E-LIS)  
 Gedung: E-LIS Subject Repository  
 Daftar Isi: "... as an Internet kiosk and as an alternative village mandi system. There are some pros and cons for this system..."

Lokasi	Repository	IOS Number
Rome	Eprints in Library and Information Science (E-LIS)	IOS2947.7210

Available online:  
[Get online](#)  
[Get online](#)

★ Favorit

---

**Fortifying Big Data infrastructures to Face Security and Privacy Issues**

Article
Info
application/pdf
eJournal

oleh **Sutikno, Toie; Universitas Ahmad Dahlan**  
 Daftar Isi: "...The explosion of data available on the internet is very increasing in recent years. One of the most..."

Lokasi	Repository	IOS Number
KOTA YOGYAKARTA	Bulletin of Electrical Engineering and Informatics	IOS160.article-957
KOTA YOGYAKARTA	TELKOMNIKA Telecommunication, Computing, Electronics and Control	IOS2608.article-957

★ Favorit



# E-resource Search

internet infrastructure

Semua Kolom

Cari: internet infrastructure

Koleksi Nasional E-resources

Persempit Pencarian

- Limit to articles from scholarly journals
- Limit to articles with full text available
- Exclude newspaper articles
- Include articles at other libraries
- Expand results using synonyms

Hapus Filter

Format: Dissertation

Format

Dissertation

Subject Area

business	1,725
economics	1,555
political science	1,345
computer science	1,261

Menampilkan 1 - 20 of 24,753 untuk pencarian: 'internet infrastructure', lama mencari: 3.13s

Sortir Relevance

**Where in the World is the Internet? Locating Political Power in Internet Infrastructure**

oleh Mathew, Ashwin Jacob

Terbitan: ProQuest Dissertations Publishing, 01-01-2014

"... My examination of Internet infrastructure focuses on the system of interconnections amongst the networks which make up the Internet, which is called..."

Available online:

[Akses full teks](#)

★ Favorit

**Planning for Broadband: An Evaluation of California's High Speed Internet Infrastructure Programs**

oleh Miller, Ryan G

Terbitan: ProQuest Dissertations Publishing, 01-01-2014

"...", which expands broadband Internet service to high-need communities in rural California..."

Available online:

[Akses full teks](#)

★ Favorit

**Policies, Pipes, and Publics: The Politics of Net Neutrality Discourse and the Regulation of Internet Infrastructures**

oleh Kimball, Danny

Terbitan: ProQuest Dissertations Publishing, 01-01-2014

"...This dissertation is a critical cultural history of net neutrality (the standard that all users and uses of the internet should receive equal treatment from network operators..."

Available online:

[Akses full teks](#)

★ Favorit



Electronic Resources

# Referral dari IOS ke Situs Partner

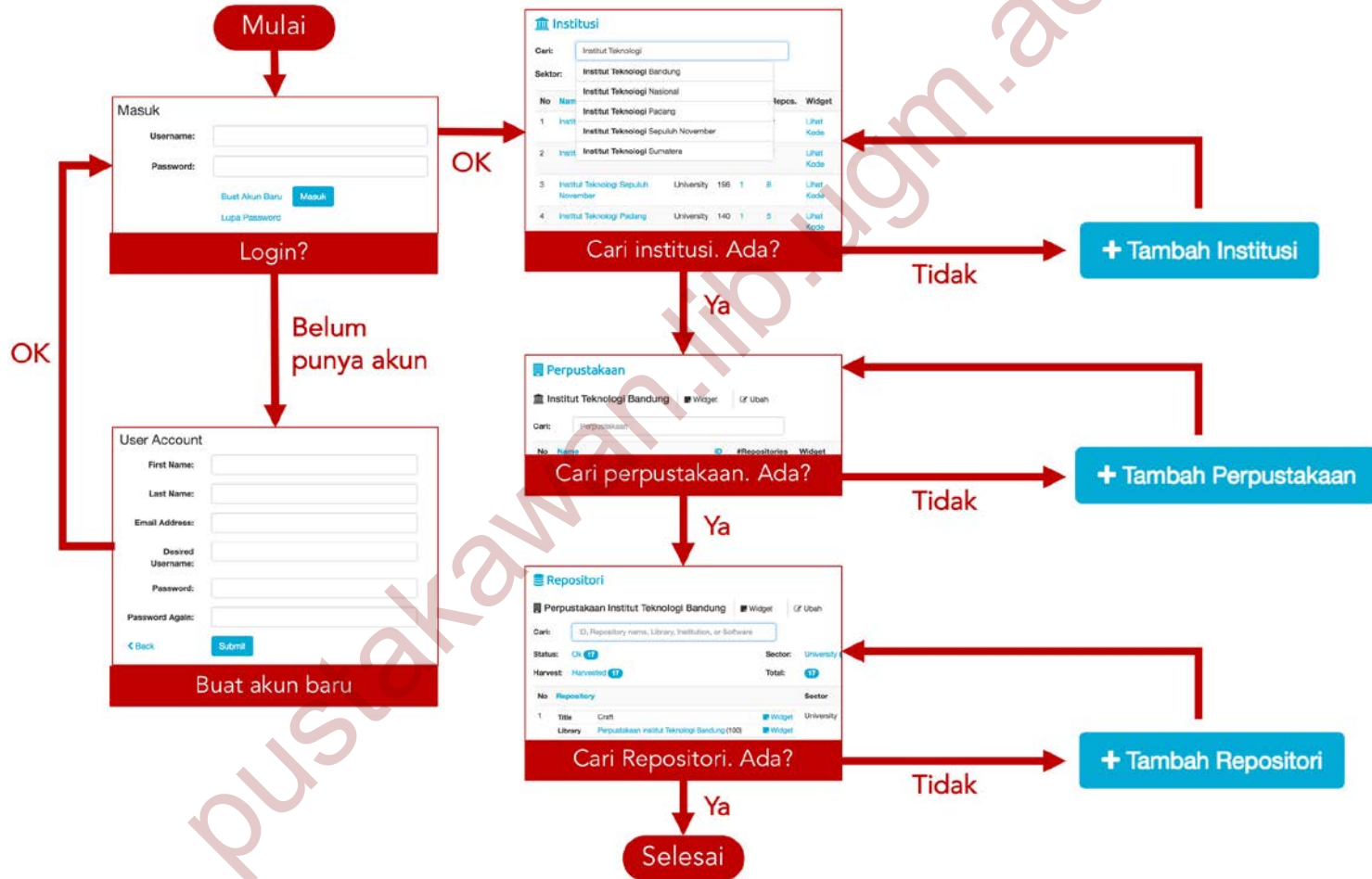
Contoh: \*.unsyah.ac.id

1 Nov 2015 – 1 Nov 2016

Pageviews by source	
Filter: Default channel grouping = Referral	
Source	Pageviews
Total	126,085 +102,727 (+439.79%)
googleweblight.com	25,174 +24,033 (+2,106.31%)
scholar.google.co.id	22,192 +17,792 (+404.36%)
int.search.tb.ask.com	21,835 +17,773 (+437.54%)
onesearch.id	11,913 +11,913 (+∞)
mysearch.com	6,433 +4,311 (+203.16%)
cobasaja.tashare.web.id	5,322 +5,322 (+∞)
ibnads.xl.co.id	3,394 +3,353 (+8,178.05%)
id.search.yahoo.com	2,317 +2,229 (+2,532.95%)

Nov 13 - Nov 19	
vs. Nov 6 - Nov 12	
Pageviews by source	
Filter: Default channel grouping = Referral	
Source	Pageviews
Total	2,253 -865 (-27.74%)
onesearch.id	462 +43 (+10.26%)
scholar.google.co.id	362 -482 (-57.11%)
googleweblight.com	347 -434 (-55.57%)
int.search.tb.ask.com	208 -61 (-22.68%)

# Cara Mendaftarkan Repositori



## Institusi

Perpustakaan

Repositori

+ Cara Daftar Repositori

Format



Struktur Informasi

## Institusi

Cari: [+ Tambah Institusi](#)Sektor: [University 364](#) [Special 41](#) [Public 28](#) [School 6](#) [National 1](#)Total: [440](#)

No	Name	Sector	ID	#Lib.	#Repos.	Widget	Updated	Action
1	<a href="#">Politeknik Kesehatan Banten</a>	University	602	1	0	<a href="#">Lihat Kode</a>	3 jam politeknik	<a href="#">Ubah</a> <a href="#">+ Perpustakaan</a> <a href="#">Q</a>
2	<a href="#">STIT Muhammadiyah Bangil Pasuruan</a>	University	601	0	0	<a href="#">Lihat Kode</a>	1 hari JIE	<a href="#">Ubah</a> <a href="#">+ Perpustakaan</a> <a href="#">Q</a>
3	<a href="#">Fakultas Hukum Universitas Padjadjaran</a>	University	600	0	0	<a href="#">Lihat Kode</a>	1 hari jbmh	<a href="#">Ubah</a> <a href="#">+ Perpustakaan</a> <a href="#">Q</a>
4	<a href="#">Indonesian Institute for Counseling, Education and Therapy (IICET)</a>	Special	599	2	0	<a href="#">Lihat Kode</a>	1 hari ifdil	<a href="#">Ubah</a> <a href="#">+ Perpustakaan</a> <a href="#">Q</a>
5	<a href="#">Fakultas Ekonomi dan Bisnis Islam Imam Bonjol</a>	University	598	0	0	<a href="#">Lihat Kode</a>	1 hari FEBI	<a href="#">Ubah</a> <a href="#">+ Perpustakaan</a> <a href="#">Q</a>
6	<a href="#">STIE Indonesia Banking School</a>	University	593	1	2	<a href="#">Lihat Kode</a>	1 hari didik_pribadi	<a href="#">Ubah</a> <a href="#">+ Perpustakaan</a> <a href="#">Q</a>
7	<a href="#">Fakultas Kesehatan Masyarakat Unhas</a>	Special	597	0	0	<a href="#">Lihat Kode</a>	1 hari Achmad Adil	<a href="#">Ubah</a> <a href="#">+ Perpustakaan</a> <a href="#">Q</a>

# Harvest Status

No	Repository	Sector	ID	Status	Updated	Action
1	<b>Title</b> Perpustakaan FITK UIN Syarif Hidayatullah Jakarta  <b>Library</b> <a href="#">Perpustakaan UIN Syarif Hidayatullah Jakarta (459)</a>  <b>Institution</b> <a href="#">UIN Syarif Hidayatullah Jakarta (394)</a>  <b>City</b> KOTA TANGERANG SELATAN <b>Homepage</b> <a href="#">Open the link</a> <b>OAI Base</b> <a href="http://opac.fitk.uinjkt.ac.id/oai2.php">http://opac.fitk.uinjkt.ac.id/oai2.php</a> <b>Prefix</b> <a href="#">oai_dc (ListRecords)</a> <b>Format</b> Multiple Types <b>Software</b> SLIMS <b>Subjects</b> Agama; Bahasa; Pendidikan <b>Groups</b> <b>IOS Page</b> <a href="#">IOS3354</a>	University	3354	<span>Ok</span> <span>Reindexed</span>	14 hari fahmi	<input type="button" value="Edit"/>

**Review:** OAI is valid

**Notes:** OK

[oai] 38 hari: Failed to parse XML response

[oai][last request] 38 hari: [http://opac.fitk.uinjkt.ac.id/oai2.php?resumptionToken=17500\\_\\_\\_oai\\_dc&verb=ListRecords](http://opac.fitk.uinjkt.ac.id/oai2.php?resumptionToken=17500___oai_dc&verb=ListRecords)

[dedup] 14 hari: 17472 processed, 4620 deduplicated

[solr] 14 hari: 12852 indexed, 0 deleted, 0 merged

## View Log Details

Repository: Perpustakaan FITK UIN Syarif Hidayatullah Jakarta (3354)

Date	User	Process	Level	Message
14 hari	scheduler	job	INFO	Finished reindex.pl
14 hari	scheduler	updateSolr	INFO	12852 indexed, 0 deleted, 0 merged
14 hari	scheduler	deduplicate	INFO	17472 processed, 4620 deduplicated
14 hari	scheduler	renormalize	INFO	34979 renormalized
14 hari	scheduler	job	INFO	Executing reindex.pl
14 hari	fahmi	deduplicate	INFO	792 processed, 785 deduplicated
38 hari	scheduler	updateSolr	INFO	14392 indexed, 0 deleted, 0 merged
38 hari	scheduler	deduplicate	INFO	17500 processed, 3108 deduplicated
38 hari	scheduler	HarvestOaiPmh	FATAL	Failed to parse XML response URL: <a href="http://opac.fitk.uinjkt.ac.id/oai2.php?resumptionToken=17500___oai_dc&amp;verb=ListRecords">http://opac.fitk.uinjkt.ac.id/oai2.php?resumptionToken=17500___oai_dc&amp;verb=ListRecords</a>
38 hari	scheduler	HarvestOaiPmh	FATAL	Error 4 at 1:22: Start tag expected, '<' not found URL: <a href="http://opac.fitk.uinjkt.ac.id/oai2.php?resumptionToken=17500___oai_dc&amp;verb=ListRecords">http://opac.fitk.uinjkt.ac.id/oai2.php?resumptionToken=17500___oai_dc&amp;verb=ListRecords</a>
38 hari	scheduler	HarvestOaiPmh	FATAL	Error 4 at 1:22: Start tag expected, '<' not found URL: <a href="http://opac.fitk.uinjkt.ac.id/oai2.php?resumptionToken=17500___oai_dc&amp;verb=ListRecords">http://opac.fitk.uinjkt.ac.id/oai2.php?resumptionToken=17500___oai_dc&amp;verb=ListRecords</a>
38 hari	scheduler	HarvestOaiPmh	FATAL	Error 4 at 1:22: Start tag expected, '<' not found

# Widget IOS

Top / UIN Syarif Hidayatullah Jakarta / Perpustakaan UIN Syarif Hidayatullah Jakarta / **Repositori**

**Koleksi Nasional**   **E-resources**

Institusi

Perpustakaan

**Repositori**

+ Cara Daftar Repositori

**Repositori**

**Perpustakaan UIN Syarif Hidayatullah Jakarta**   **Widget**

**Cari:**

→ ↻ [onesearch.id/Repositories/Widget?library\\_id=459](https://onesearch.id/Repositories/Widget?library_id=459) ☆ ⚙

## Perpustakaan UIN Syarif Hidayatullah Jakarta

### Contoh Kode Widget

#### Kode Widget

Salin (copy) kode di bawah ini, lalu tempel (paste) di dalam tag <body> halaman situs anda, dimana anda ingin form pencarian dan hasilnya ditampilkan.

```
<div id="ios-widget"></div>
<script>
(function() {
  var id = '459';
  var fr = document.createElement('iframe');
  fr.style.width = "100%";
  fr.style.height = "700px";
  fr.src = "http://onesearch.id/Search/Results?widget=1&library_id="+id;
  var x = document.getElementById("ios-widget");
  x.appendChild(fr);
})();
</script>
```

# Widget: OneSearch Institusi

Indonesia OneSearch

indonesia **oneSearch** by PERPUSNAS

Perpustakaan UIN Syarif Hidayatullah Jakarta  
KOTA TANGERANG SELATAN, BANTEN

Cari buku dan e-resources Semua Kolom Q Cari

Cari:

**Persempit Pencarian**

Hapus Filter

Other: Perpustakaan UIN Syarif Hidayatullah Jakarta

**Koleksi**

Perpustakaan FITK UIN Syarif Hidayatullah Jakarta	17,486
Katalog Fakultas Ushuluddin UIN Syarif Hidayatullah Jakarta	7,299
MANUSKRIPTA	1,733
STUDIA ISLAMIKA	490
AL-IQTISHAD	174

Menampilkan 1 - 20 of 28,654 untuk pencarian: "", lama mencari: 0.48s

**Logika Praktis**  
Book  
oleh BAKRY, Noor MS  
Terbitan: Liberty, 1986  
Institusi: UIN Syarif Hidayatullah Jakarta  
Gedung: Perpustakaan UIN Syarif Hidayatullah Jakarta

Lokasi	Repository	IOS Number
KOTA TANGERANG SELATAN	Katalog Fakultas Ushuluddin UIN Syarif Hidayatullah Jakarta	<a href="#">IOS3389.slims-6053</a>

Available online:  
[Get online](#)

**Kebatinan dan Injil**  
Book  
oleh Hadiwijono, Harun  
Terbitan: BPK Gunung Mulia, 2002  
Institusi: UIN Syarif Hidayatullah Jakarta  
Gedung: Perpustakaan UIN Syarif Hidayatullah Jakarta

# Knowledge Graph

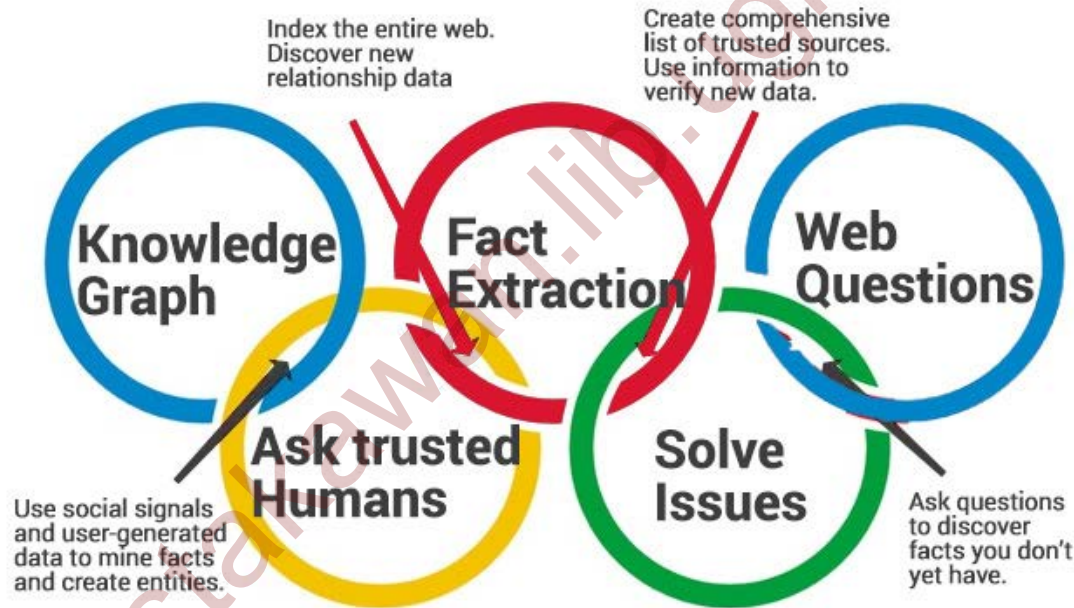
Text Analysis

pustakawen.lib.ugm.ac.id



# Google Knowledge Graph

## The 5 Steps of Google's Semantic Search



**Each step leads to a more semantic web**

# Contoh: Semantic Search

Google how old is ismail fahmi

About 497,000 results (0.60 seconds)

**Ismail Fahmi / Age at death**

**75**  
1922–1997

**Ismail Fahmi**  
Diplomat

Ismail Fahmi was an Egyptian diplomat and politician. He served as ambassador to Austria, tourism minister, foreign minister and deputy prime minister. He was awarded to professorship. [Wikipedia](#)

**Born:** 1922, Egypt  
**Died:** 1997, Cairo, Egypt  
**Children:** Nabil Fahmy  
**Books:** [Negotiating for peace in the Middle East](#)  
**People also search for:** [Nabil Fahmy](#), [Harry Hunkele](#), [Jonathan P. Hicks](#), [Arick Wierson](#), [Matthew Tollin](#)

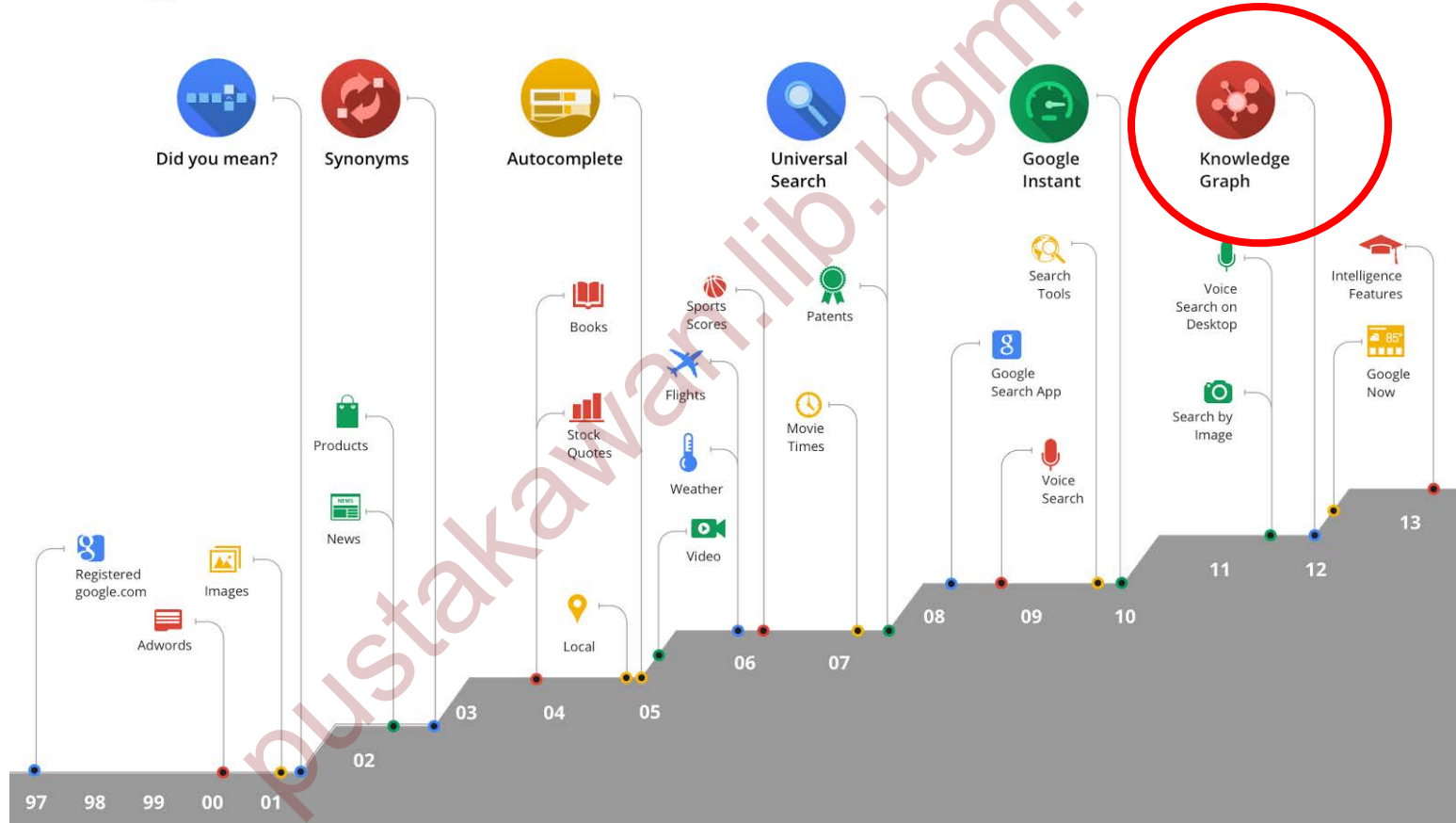
Mengenal sosok Ismail Fahmi PhD Nara Sumber SLIMS COMMEET ...  
[www.slimscommeet.web.id/.../134-mengenal-sosok-ismail-fahmi-p...](http://www.slimscommeet.web.id/.../134-mengenal-sosok-ismail-fahmi-p...) Translate this page  
Jul 11, 2015 - Ismail Fahmi merupakan sarjana elektro lulusan Institut Teknologi Bandung (ITB) dan mengambil tingkat pascasarjananya di Universitas ...

**Ismail Fahmi Profiles | Facebook**  
<https://www.facebook.com/public/Ismail-Fahmi>  
View the profiles of people named Ismail Fahmi. Join Facebook to connect with Ismail Fahmi and others you may know. Facebook gives people the power to...

**Fact Extraction**

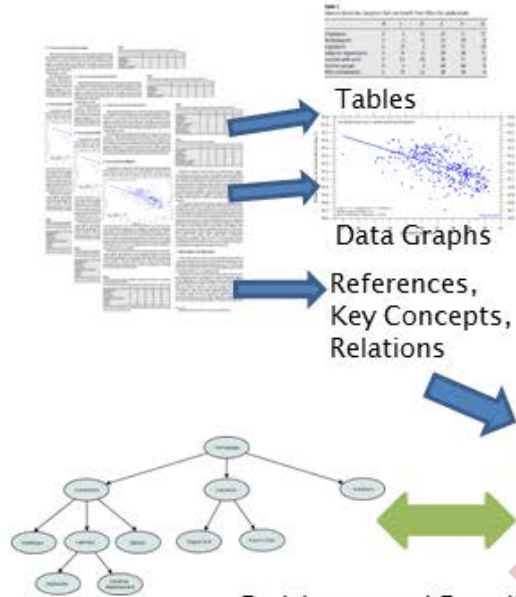
# Jalan Panjang Google

## Google Search Timeline



# Building Knowledge Graph

Unstructured/Semistructured content



Relational DB



External Domain Data

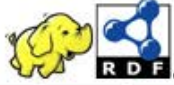
← Align →



Customer Data



Knowledge Graph

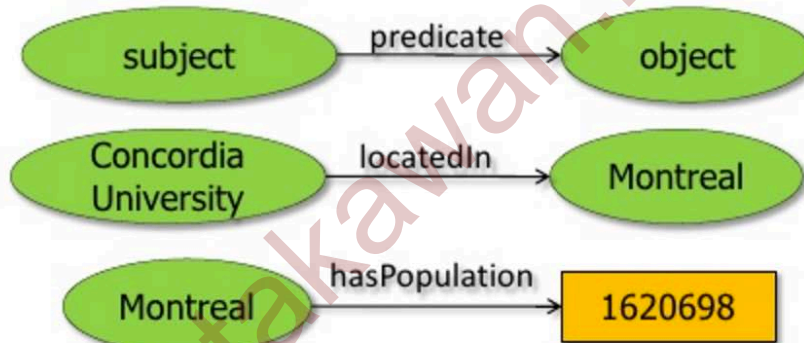


- Search++
- Recommendations
- Vertical applications
- Explorative interfaces

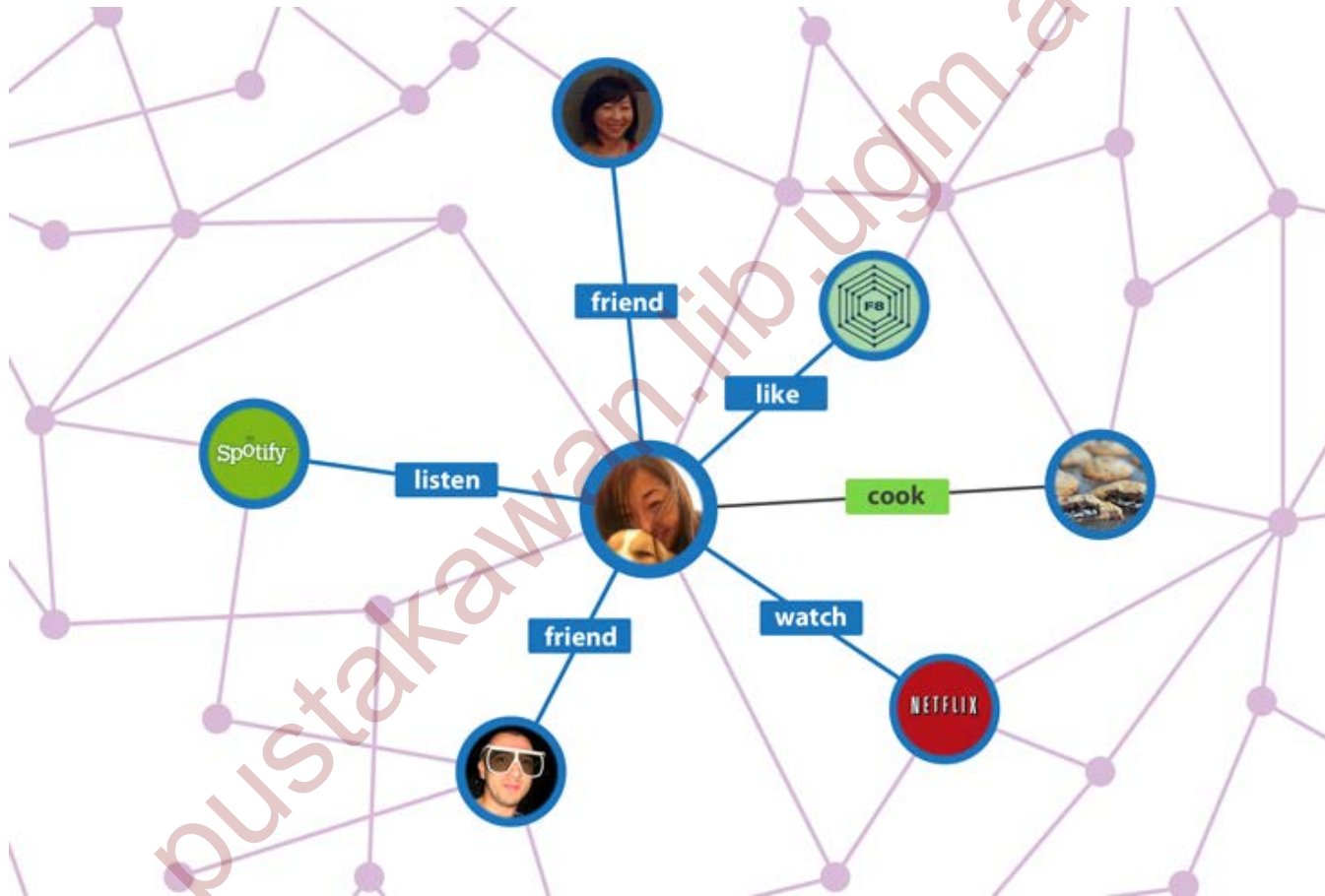
pustakawan.lib.ugm.ac.id

# Everything is a Triple (RDF)

- Everything is a triple
  - **Subject** (resource), **Predicate** (relation), **Object** (resource or literal)
- The RDF graph is a collection of triples

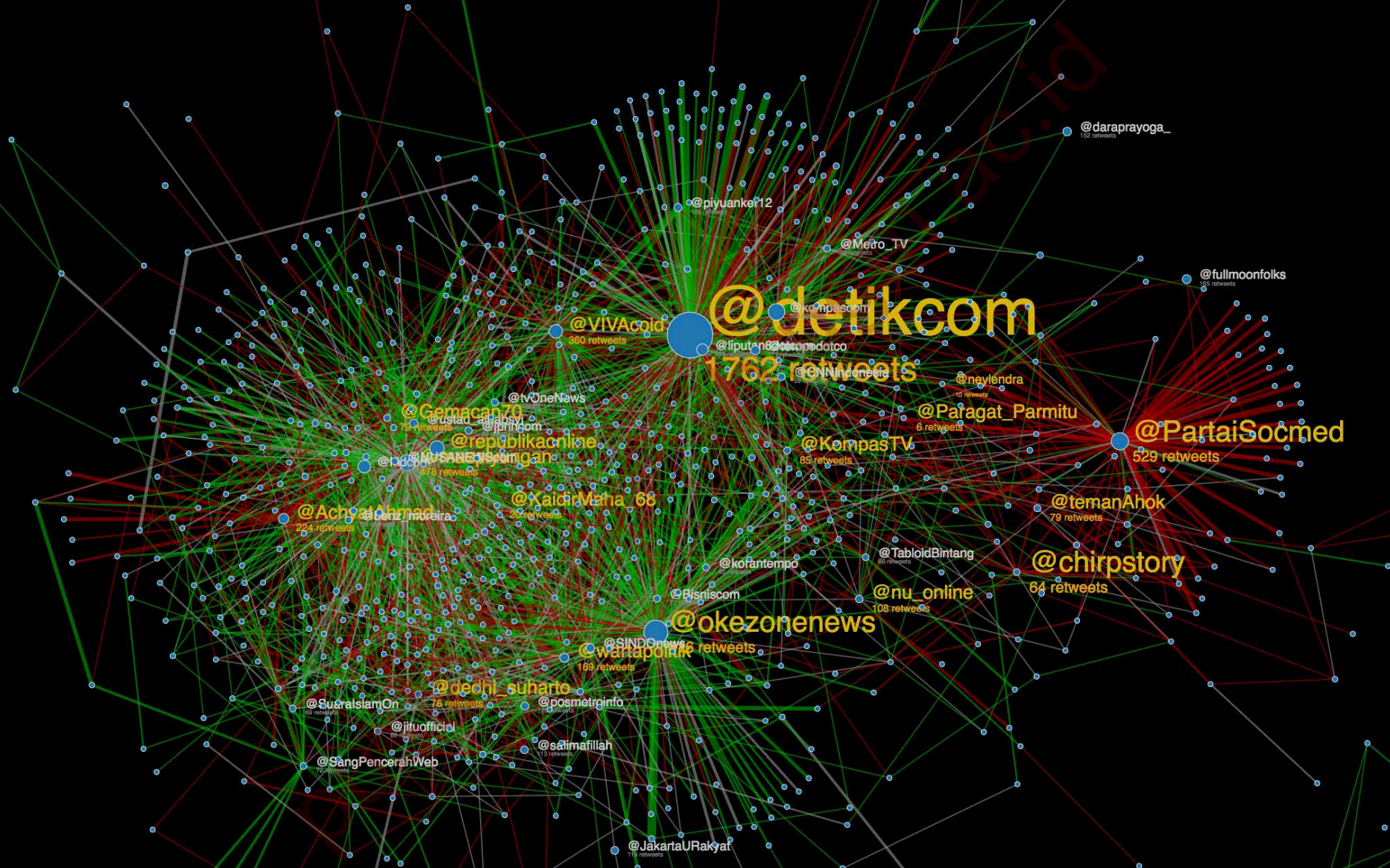


# Facebook Knowledge Graph



# Twitter Knowledge Graph





**@detikcom**  
1762 retweets

**@VIVAcoi**  
360 retweets

**@PartaiSocmed**  
529 retweets

**@republikonline**

**@KompasTV**  
85 retweets

**@Achy**  
224 retweets

**@Xaichyana\_68**  
20 retweets

**@temanAhok**  
79 retweets

**@okezonenews**  
108 retweets

**@chirpstory**  
64 retweets

**@wartabinc**  
169 retweets

**@dechi\_suharto**  
76 retweets

**@SuaraIslamOn**  
60 retweets

**@posmetroinfo**  
10 retweets

**@SangPencerahWeb**  
7 retweets

**@satimafillah**  
13 retweets

**@JakartaURakyat**  
118 retweets

**@daraprayoga\_**  
152 retweets

**@piyuanke12**  
194 retweets

**@Metro\_TV**  
10 retweets

**@fullmoonfolks**  
163 retweets

**@kompascom**

**@liputan6dotcom**

**@CNNIndonesia**

**@nevindra**  
10 retweets

**@Paragat\_Pamitu**  
6 retweets

**@tvOneNews**

**@Gemaca70**  
79 retweets

**@ustad\_silabsy**

**@jprn.com**

**@MUSANEVS**  
178 retweets

**@Sanzmora**

**@TabloidBintang**  
6 retweets

**@korantempo**

**@Pisniscam**

**@nu\_online**  
108 retweets

**@SINDOnews**

**@jituofficial**  
6 retweets

**@SangPencerahWeb**  
7 retweets



# Natural Language Processing

Introduction

pustakawarnitib.ugm.ac.id

# Teknologi NLP

---

- Segmentation
- Part-Of-Speech (POS) Tagging
- Automatic Term Recognition (ATR)
- Named Entity Recognition (NER)
- Term Relationship Extraction
- Syntax Analysis (S-P-O)

pustakawan.lib.ugm.ac.id

# Unstructured Text

KANZ PHILOSOPHIA, Volume 3, Number 2, December 2013 201

## KEWALIAN DALAM TASAWUF NUSANTARA

Artikel ini mendiskusikan doktrin Kewalian (al-walāyah) yang mempunyai basis yang kuat dalam al-Qur'an dan hadis dan isu ini secara sistematis dibahas melalui ajaran Ibn 'Arabi yang kompleks.

Sementara dalam kajian tasawuf di Nusantara, sebutan wali ini sudah mulai dikenal bersamaan dengan masuknya Islam ke negeri ini yang mengacu kepada dua pengertian, wali sebagai orang yang memiliki kesaktian-kesaktian (occulties) sebagai implikasi dari kekeramatan dan wali sebagai penguasa wilayah tertentu. Terlepas dari pengertian dan cakupannya, ajaran kewalian yang sesungguhnya tidak lepas dari ajaran tentang kenabian dan kerasulan yang menyiratkan pesan bahwa dunia manusia bukan hanya dunia material yang identik dengan kenikmatan hedonis, tetapi di balik dunia fi sik terdapat dunia metafisik yang belum banyak diketahui manusia dan dari sanalah dunia fisik ini dikendalikan.

Kata-kata Kunci : wali, karamah, Nur Muhammad, kewalian, kenabian, penutup para wali  
Fakultas Ushuluddin, UIN Syarif Hidayatullah. E-mail : yunasrilali@gmail.com

# Segmentation

=====page1=====

-----par-----

KANZ PHILOSOPHIA, Volume 3, Number 2, December 2013 201

-----par-----

KEWALIAN DALAM TASAWUF NUSANTARA

-----par-----

Artikel ini mendiskusikan doktrin Kewalian (al-walāyah) yang mempunyai basis yang kuat dalam al-Qur'an dan hadis dan isu ini secara sistematis dibahas melalui ajaran Ibn 'Arabi yang kompleks.

-----par-----

Sementara dalam kajian tasawuf di Nusantara, sebutan wali ini sudah mulai dikenal bersamaan dengan masuknya Islam ke negeri ini yang mengacu kepada dua pengertian, wali sebagai orang yang memiliki kesaktian-kesaktian (occulties) sebagai implikasi dari kekeramatan dan wali sebagai penguasa wilayah tertentu.

Terlepas dari pengertian dan cakupannya, ajaran kewalian yang sesungguhnya tidak lepas dari ajaran tentang kenabian dan kerasulan yang menyiratkan pesan bahwa dunia manusia bukan hanya dunia material yang identik dengan kenikmatan hedonis, tetapi di balik dunia fi sik terdapat dunia metafisik yang belum banyak diketahui manusia dan dari sanalah dunia fi sik ini dikendalikan.

-----par-----

Kata-kata Kunci : wali, karamah, Nur Muhammad, kewalian, kenabian, penutup para wali 1 Fakultas Ushuluddin, UIN Syarif Hidayatullah. E-mail : yunasrilali@ymail.com

# Part-of-Speech Tagging

----- par -----

Sementara\_SC dalam\_IN kajian\_NN tasawuf\_VB di\_IN Nusantara\_NNP ,\_Z sebutan\_NN wali\_NN ini\_PR sudah\_MD mulai\_VB dikenal\_VB bersamaan\_VB dengan\_IN masuknya\_X Islam\_NNP ke\_IN negeri\_NN ini\_PR yang\_SC mengacu\_VB kepada\_IN dua\_CD pengertian\_NN ,\_Z wali\_NN sebagai\_IN orang\_NN yang\_SC memiliki\_VB kesaktian-kesaktian\_NN -LRB\_-LRB- occulties\_FW -RRB\_-RRB- sebagai\_IN implikasi\_NN dari\_IN kekeramatan\_NN dan\_CC wali\_NN sebagai\_IN penguasa\_NN wilayah\_NN tertentu\_PR .\_Z

Terlepas\_VB dari\_IN pengertian\_NN dan\_CC cakupannya\_NN ,\_Z ajaran\_NN kewalian\_NN yang\_SC sesungguhnya\_RB tidak\_NEG lepas\_VB dari\_IN ajaran\_NN tentang\_IN kenabian\_NN dan\_CC kerasulan\_NN yang\_SC menyiratkan\_VB pesan\_NN bahwa\_SC dunia\_NN manusia\_NN bukan\_NEG hanya\_RB dunia\_NN material\_NN yang\_SC identik\_JJ dengan\_SC kenikmatan\_NN hedonis\_JJ ,\_Z tetapi\_CC di\_IN balik\_NN dunia\_NN fi\_X sik\_X terdapat\_VB dunia\_NN metafisik\_VB sik\_X yang\_SC belum\_NEG banyak\_CD diketahui\_VB manusia\_NN dan\_CC dari\_IN sanalah\_NN dunia\_NN fi\_X sik\_X ini\_PR dikendalikan\_VB .\_Z

# Terms & Named Entities

```
{
  "lang": "id",
  "who_str_mv": [
    "George Edward Moore (PER)",
    "Bertrand Russerll (PER)",
    "Vienna Circle7 (ORG)"
  ],
  "organization_ent": [
    "Vienna Circle7"
  ],
  "type": "sent",
  "where": [
    ""
  ],
  "adj": [
    "terakhir",
    "lebih radikal",
    "penting"
  ],
  "person_ent": [
    "Bertrand Russerll",
    "George Edward Moore"
  ],
  "where_str_mv": [
    "Inggris",
    "Vienna"
  ],
  "ext_id": "IOS2669.article-25",
  "source": "IOS2669",
  "how": [
    ""
  ],
  "verb": [
    "berawal",
    "menemukan",
    "lewat"
  ],
  "term": [
    "Kecenderungan",
    "Vienna Circle7",
    "bentuknya",
    "Inggris",
    "tokoh-tokoh penting",
    "Bertrand Russerll",
    "George Edward Moore"
  ],
  "page_id": "page-561c1ca02633b6809d7d6f88cb31a826-5",
  "repoId": "IOS2669",
  "institution_id": "203",
  "why": [
    ""
  ],
  "when": [
    ""
  ],
  "id": "sent-561c1ca02633b6809d7d6f88cb31a826-5-1-70",
  "create_date": "2016-09-22T15:26:15Z",
  "location_ent": [
    "Inggris"
  ],
  "par_id": "par-561c1ca02633b6809d7d6f88cb31a826-5-1",
  "doc_id": "doc-561c1ca02633b6809d7d6f88cb31a826",
  "sentence_tag": "Kecenderungan_NN terakhir_JJ ini_PR sebetulnya_RB berawal_VB dari_IN Vienna_NNP Circle7_NNP",
  "sentence_tx": "Kecenderungan terakhir ini sebetulnya berawal dari Vienna Circle7 dan menemukan bentuknya ya"
},
```

Knowledge Graph hasil ekstraksi dari text

# SPO Triples

```
"response": {
  "numFound": 858,
  "start": 0,
  "docs": [
    {
      "ext_id": "IOS2669.article-25",
      "p_long": "memberi",
      "s_long": "lsafat Islam",
      "o_long": "kedudukan yang demikian tinggi terhadap modus pengetahuan primordial seperti al -"
    },
    {
      "o_long": "kemungkinan menyatukan kedua genre fi lsafat tersebut",
      "s_long": "lsafat Islam",
      "p_long": "memandang",
      "ext_id": "IOS2669.article-25"
    },
    {
      "ext_id": "IOS2669.article-25",
      "p_long": "menunjukkan",
      "s_long": "prinsipnya pendekatan Islam",
      "o_long": "bahwa kedua sistem epistemologi yang tampaknya berlawanan itu -- Platonik dan Aristotelian -- bisa di"
    },
    {
      "p_long": "berpendirian",
      "ext_id": "IOS2669.article-25",
      "o_long": "bahwa pikiran pada",
      "s_long": "fi lsafat Islam"
    },
    {
      "s_long": "Filsafat Islam",
      "o_long": "bahwa analisis Aristotelian tentang",
      "ext_id": "IOS2669.article-25",
      "p_long": "meyakini"
    },
  ],
}
```

# Text Analysis

Explorer.OneSearch.id

pustakawan.lib.ugm.ac.id



# User Interface

Halaman muka Knowledge Explorer IOS.

- Tiap folder berisi artikel yang sudah diproses menggunakan NLP dan fact mining.
- Kolom sebelah kiri untuk pencarian dan filtering.
- Menu di sebelah atas untuk menampilkan jenis analisis.

The screenshot displays the Knowledge Explorer interface in a web browser. The browser address bar shows 'explorer.onesearch.id'. The page header includes the 'Indonesia oneSearch by PERPUSNAS' logo and navigation tabs: 'SPO Tree', 'Co-occurrence', 'Network Graph', 'Network Activity Graph', 'Concept Graph', 'Ontology Graph', and 'Knowledge Explorer'. On the left side, there is a search and filter panel with the following sections:

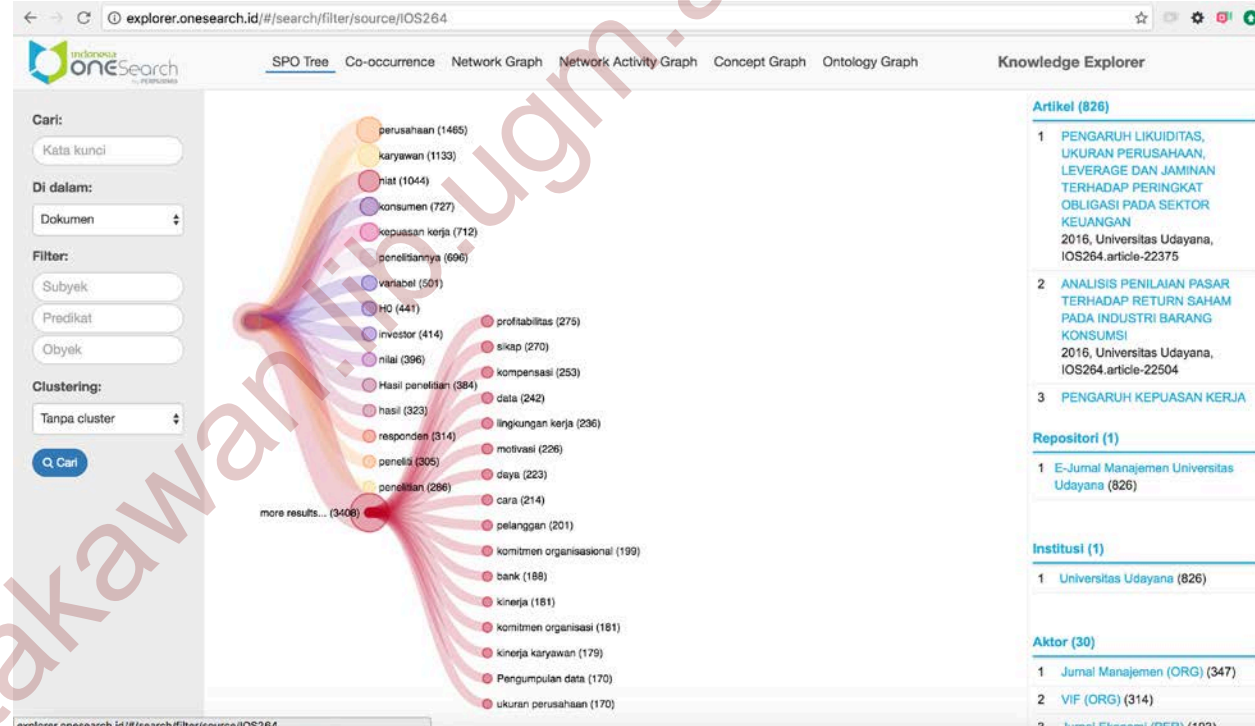
- Cari:** A search input field with the placeholder 'Kata kunci'.
- Di dalam:** A dropdown menu currently set to 'Dokumen'.
- Filter:** Three input fields for 'Subyek', 'Predikat', and 'Objek'.
- Clustering:** A dropdown menu currently set to 'Tanpa cluster'.
- A blue 'Cari' button.

The main content area is titled 'Repositori' and displays a grid of 22 folder icons, each representing a repository with a specific number of articles:

Repository ID	Number of Articles
IOS264	826
IOS14	669
IOS2852	621
IOS3233	442
IOS3457	431
IOS297	411
IOS3268	399
IOS236	219
IOS102	217
IOS915	214
IOS3344	173
IOS3246	162
IOS893	162
IOS2153	160
IOS2315	159
IOS2663	158
IOS2395	153
IOS2232	145
IOS1308	136
IOS585	118

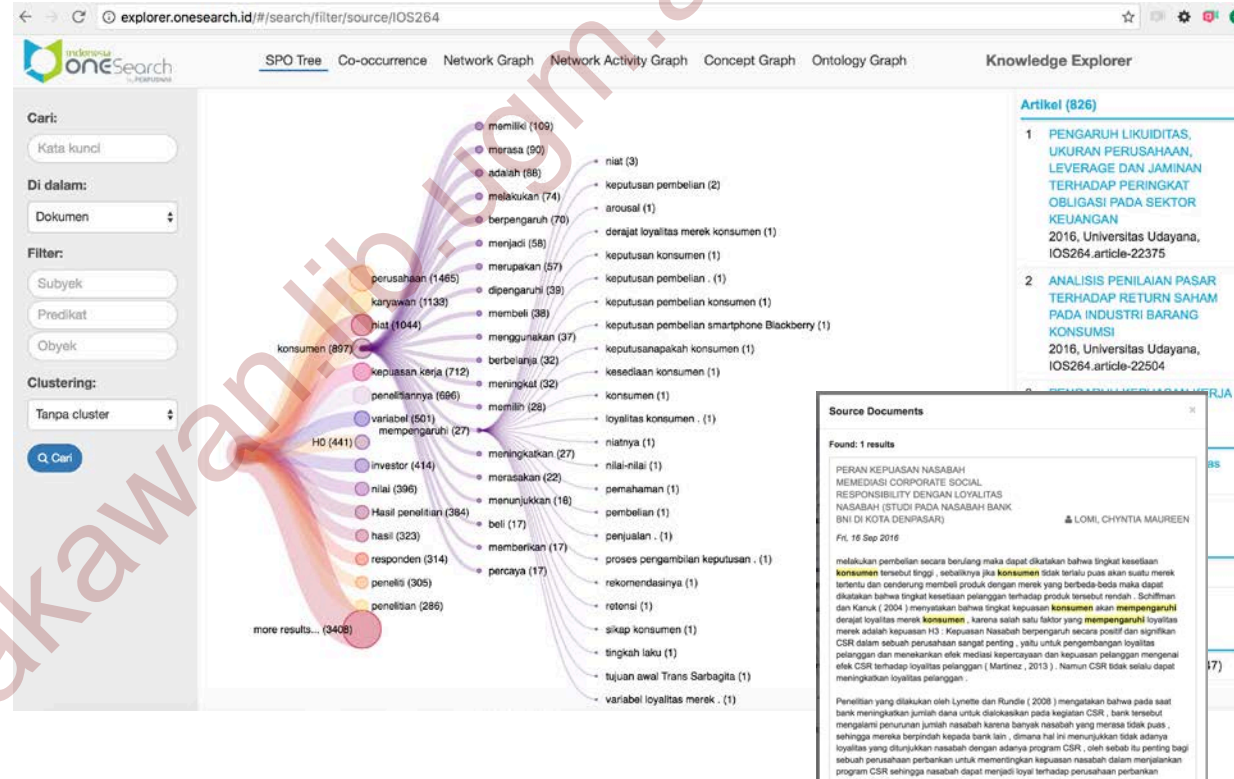
# SPO Tree Repositori

- Klik salah satu folder.
- Analisis pertama yang ditampilkan adalah SPO Tree.
- Di bagian tengah ditampilkan 30 subyek yang paling sering muncul dalam folder repositori.
- Di sisi kanan ditampilkan daftar artikel, repositori, institusi, author, dll yang bisa digunakan untuk memfilter cakupan dokumen yang akan dianalisis.



# SPO Tree: Explore

- Untuk melihat apa predikat dari sebuah subyek, klik subyek yang Anda ingin ketahui, lalu akan muncul daftar predikat.
- Dan untuk melihat daftar obyek, klik predikat, sehingga ditampilkan SPO tree seperti di samping.
- Klik obyek, maka akan tampil halaman dari dokumen yang berisi triple SPO tersebut.



**Sources Documents**

Found: 1 results

PERAN KEPUASAN NASABAH MEMEDIASI CORPORATE SOCIAL RESPONSIBILITY DENGAN LOYALITAS NASABAH (STUDI PADA NASABAH BANK BNI DI KOTA DENPASAR)

▲ LOMI, CHYNTIA MAUREEN

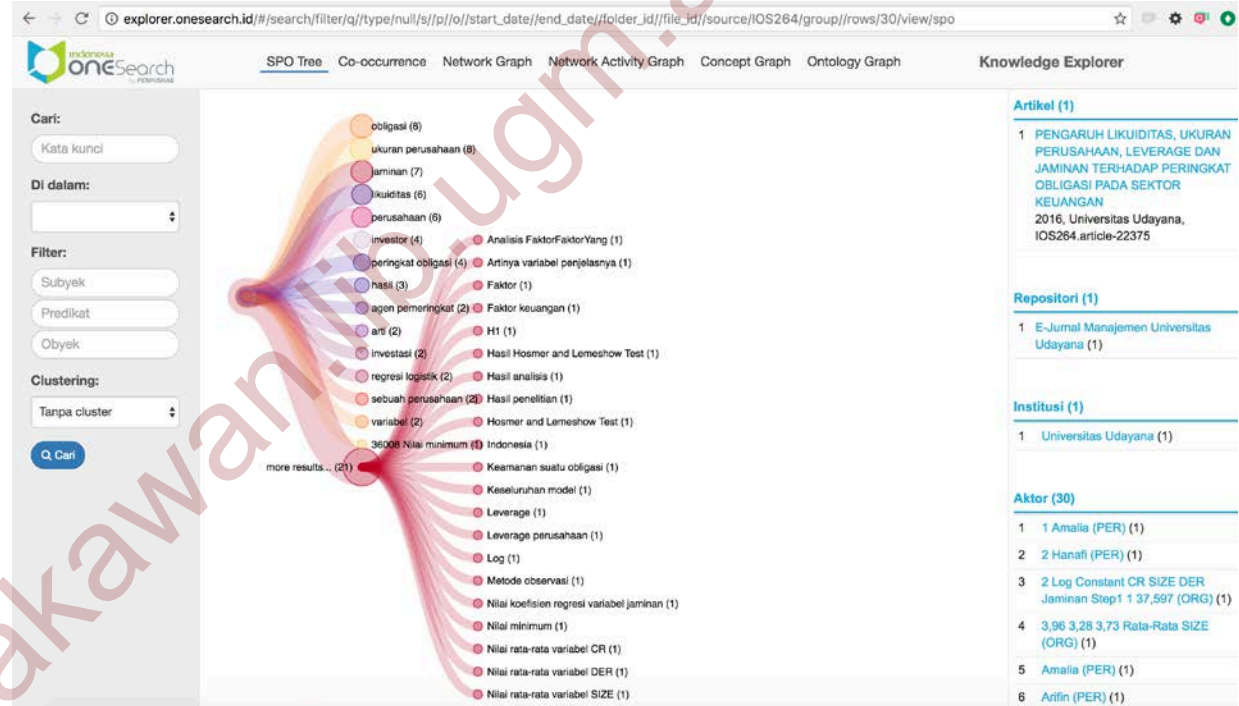
16 Sep 2016

melakukan pembelian secara berulang maka dapat dikatakan bahwa tingkat kesediaan **konsumen** tersebut tinggi; sebaliknya jika **konsumen** tidak terlalu puas akan suatu merek tertentu dan cenderung membeli produk dengan merek yang berbeda-beda maka dapat dikatakan bahwa tingkat kesediaan pelanggan terhadap produk tersebut rendah. Schiffman dan Kanuk (2004) menyatakan bahwa tingkat kepuasan **konsumen** akan **mempengaruhi** derajat loyalitas merek **konsumen**, karena salah satu faktor yang **mempengaruhi** loyalitas merek adalah kepuasan H3. Kepuasan Nasabah berpengaruh secara positif dan signifikan CSR dalam sebuah perusahaan sangat penting, yaitu untuk pengembangan loyalitas pelanggan dan menekankan efek mediasi kepercayaan dan kepuasan pelanggan mengenai efek CSR terhadap loyalitas pelanggan (Martinez, 2013). Namun CSR tidak selalu dapat meningkatkan loyalitas pelanggan.

Penelitian yang dilakukan oleh Lynette dan Rundle (2008) mengatakan bahwa saat saat bank meningkatkan jumlah dana untuk disalurkan pada kegiatan CSR, bank tersebut mengalami penurunan jumlah nasabah karena banyak nasabah yang merasa tidak puas, sehingga mereka berpindah kepada bank lain, dimana hal ini menunjukkan tidak adanya loyalitas yang ditunjukkan nasabah dengan adanya program CSR, oleh sebab itu penting bagi sebuah perusahaan perbankan untuk mementingkan kepuasan nasabah dalam menjalankan program CSR sehingga nasabah dapat menjadi loyal terhadap perusahaan perbankan

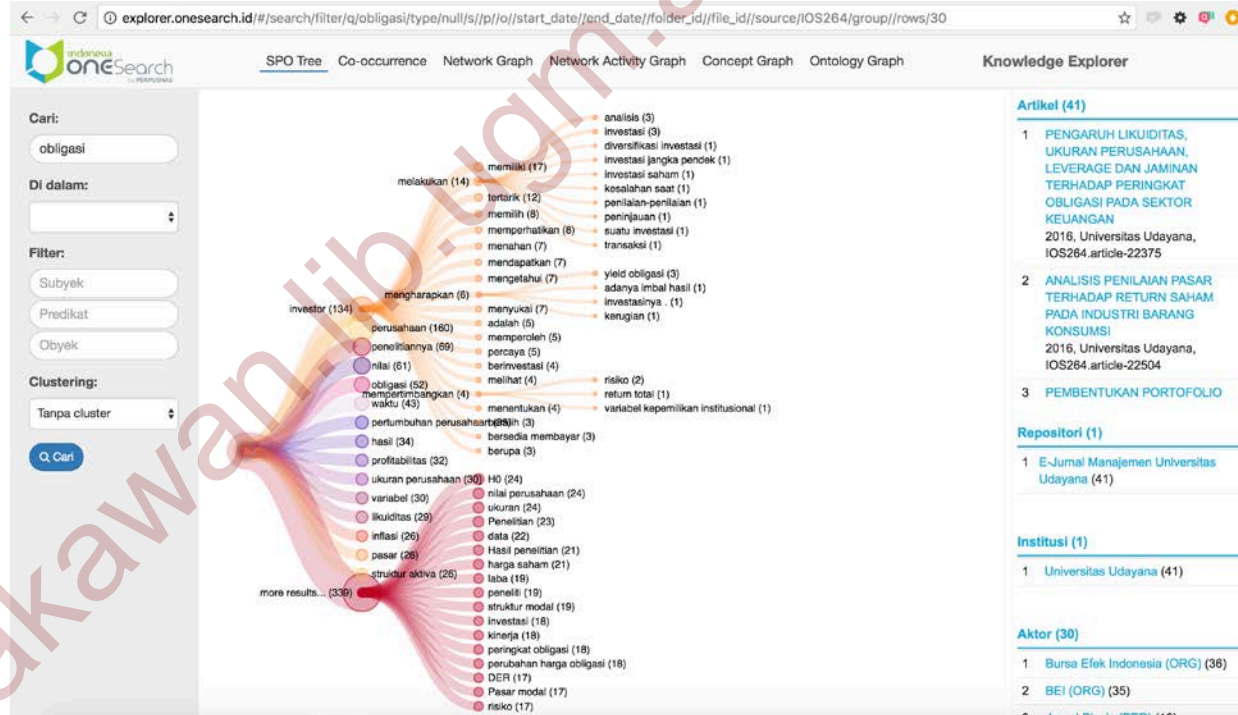
# SPO Tree Dokumen

- Untuk melihat SPO Tree dari sebuah dokumen, klik judul dokumen di kolom sebelah kanan.
- Maka grafik akan diperbarui dengan SPO yang baru.



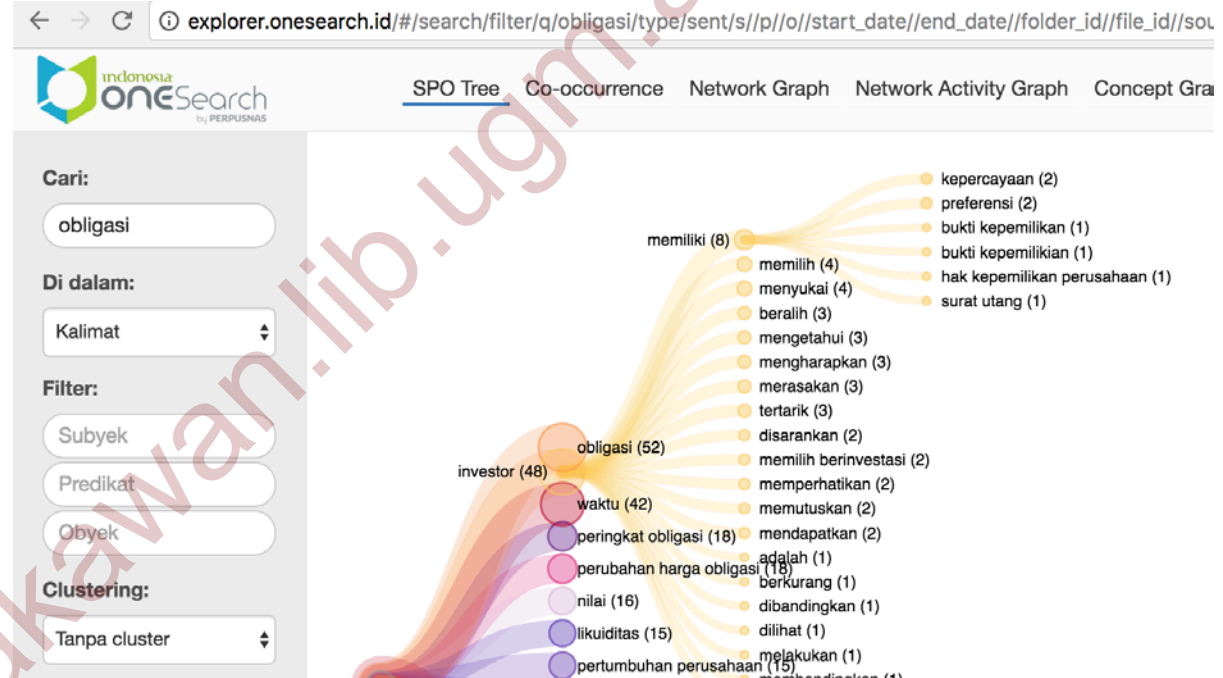
# Search

- Untuk menampilkan analisis terhadap dokumen yang mengandung kata kunci tertentu, masukkan kata kunci ke dalam kolom Cari, lalu enter.
- Maka analisis akan dibatasi pada dokumen hasil pencarian saja.



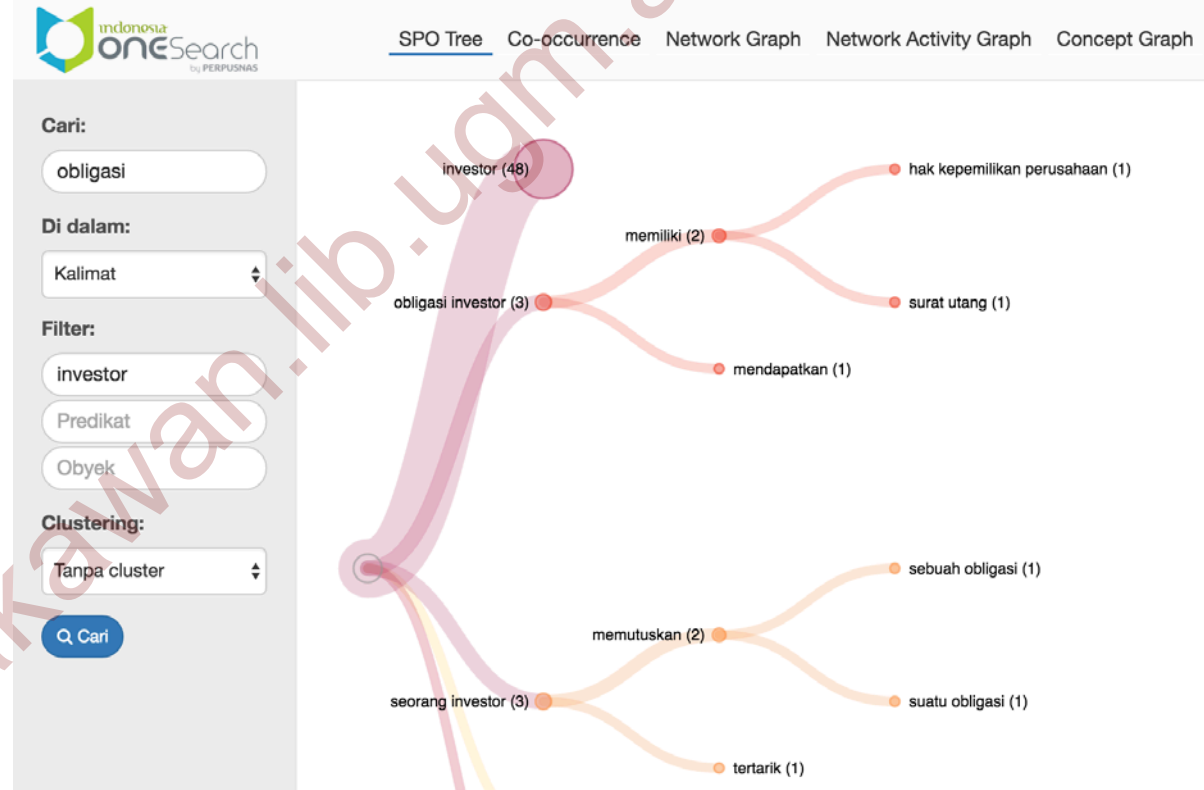
# Proximity

- Secara default, proximity atau cakupan pencarian dilakukan pada seluruh dokumen.
- Kita bisa mengubah proximity menjadi lebih ketat, dengan memilih cakupan pada kolom 'Di dalam':
  - Dokumen
  - Halaman
  - Paragraph
  - Kalimat
- Contoh di gambar kanan, untuk pencarian obligasi kita set proximity pada level 'kalimat'. Artinya, SPO yg ditampilkan hanya dari kalimat yang mengandung kata kunci saja.



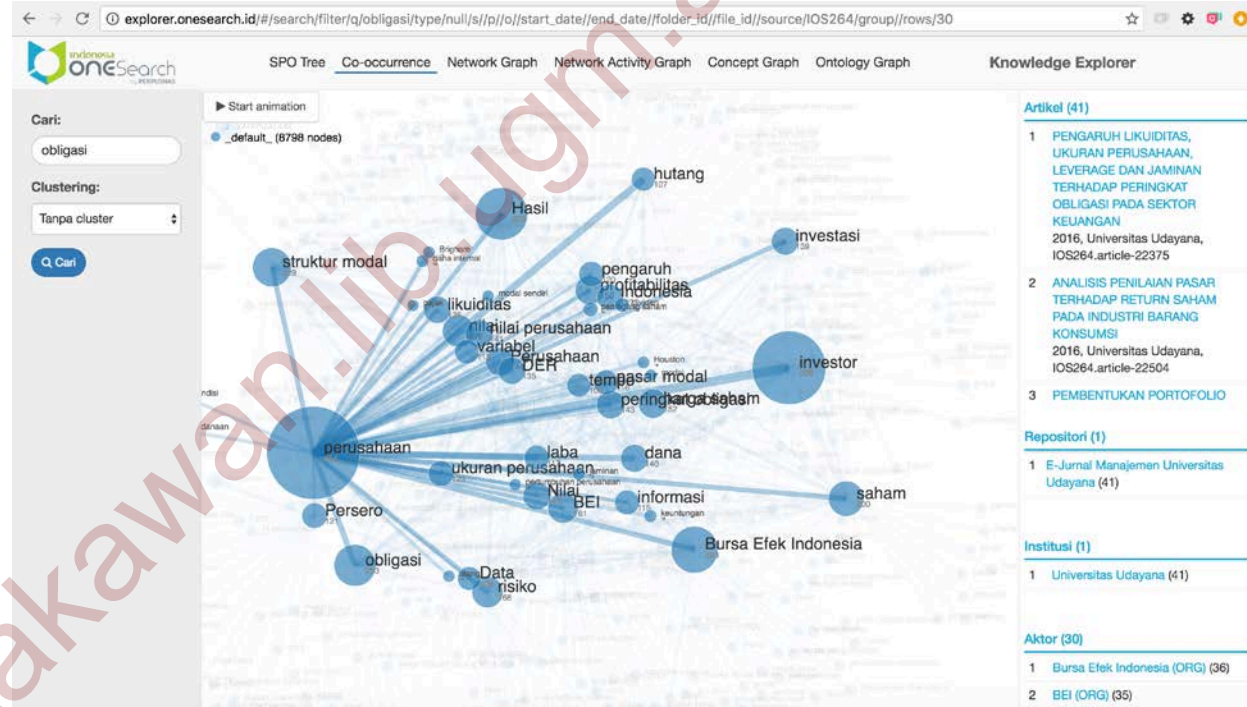
# Filter SPO

- Filter Subyek, Predikat, dan Obyek berguna untuk menampilkan relasi SPO yang mengandung kata kunci dari kolom yang terkait.
- Contoh, pada kata kunci obligasi di samping, kita ingin melihat semua subyek yang mengandung kata kunci 'investor'.



# Co-occurrence

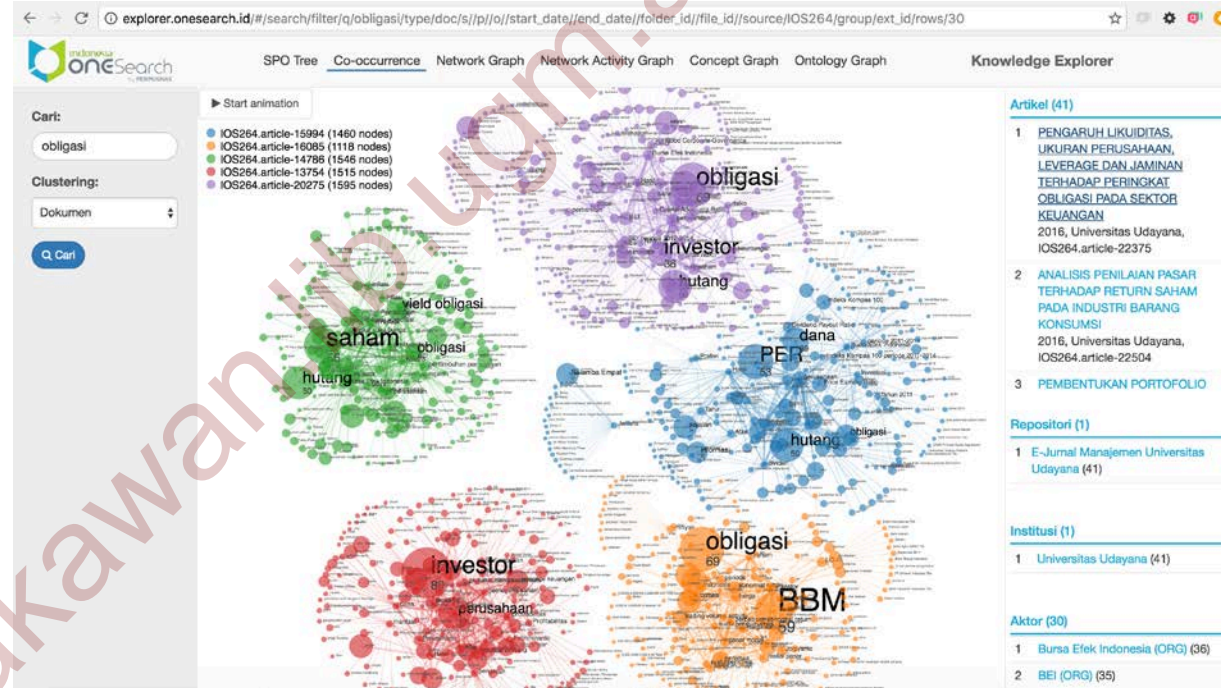
- Klik menu 'Co-occurrence', maka akan ditampilkan network graph dari konsep-konsep yang sering muncul bersamaan dalam paragraph.
- Contoh untuk kata kunci 'obligasi', gambar di samping memperlihatkan bagaimana 'perusahaan' sering ditemukan dengan konsep lain seperti Investor, struktur modal, pasar modal, dll.





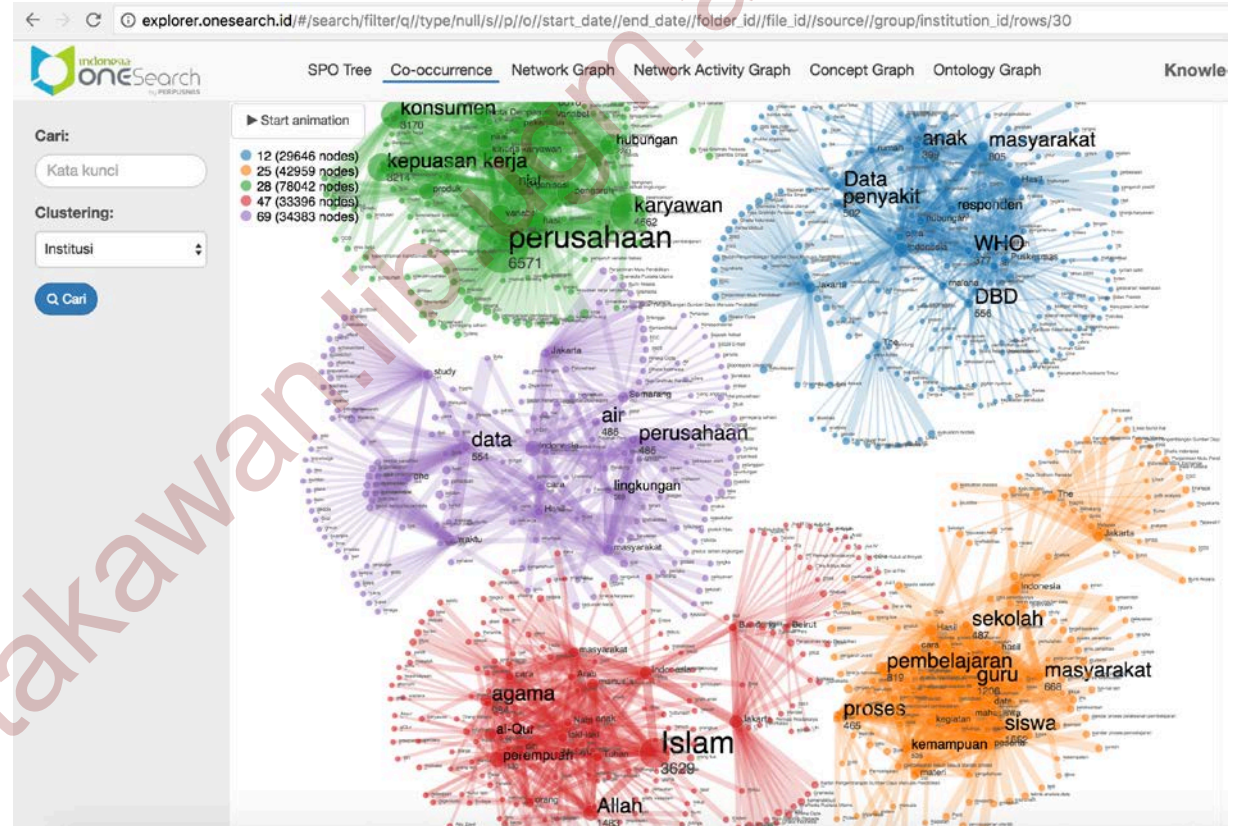
# Clustering

- Pilih mode Clustering yang ingin digunakan. Ada beberapa pilihan:
  - Tanpa cluster
  - Institusi
  - Repositori
  - Dokumen
- Dalam contoh di samping kita mengelompokkan cooccurrence untuk kata kunci 'obligasi' berdasarkan Dokumen. Maksimal kita bisa lihat 5 item dalam satu perbandingan.
- Dari cluster ini, kita bisa lihat dan bandingkan konsep2 utama yang sering muncul bersamaan dalam dokumen.



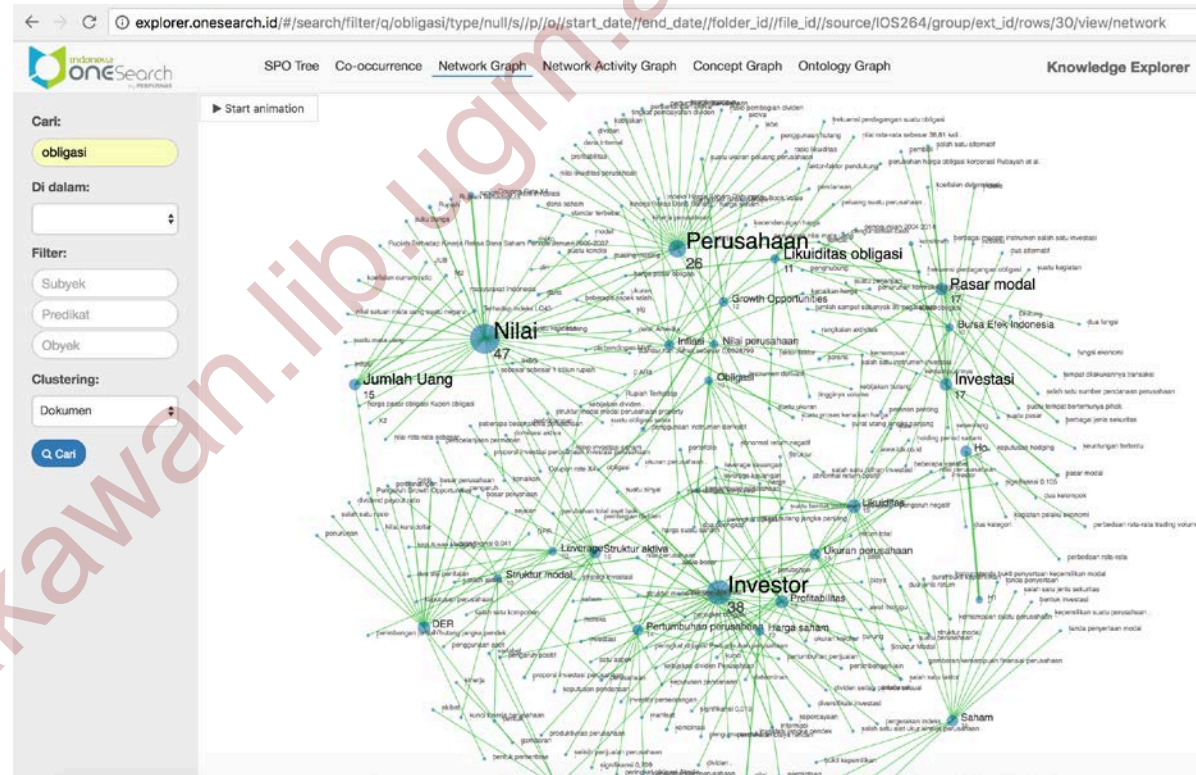
# Peta Profile Institusi

Ketika pada mode clustering kita pilih Institution, maka perbandingan peta knowledge akan ditampilkan dan dikelompokkan berdasarkan topik yang paling dominan dan yang saling terkait melalui cooccurrence.



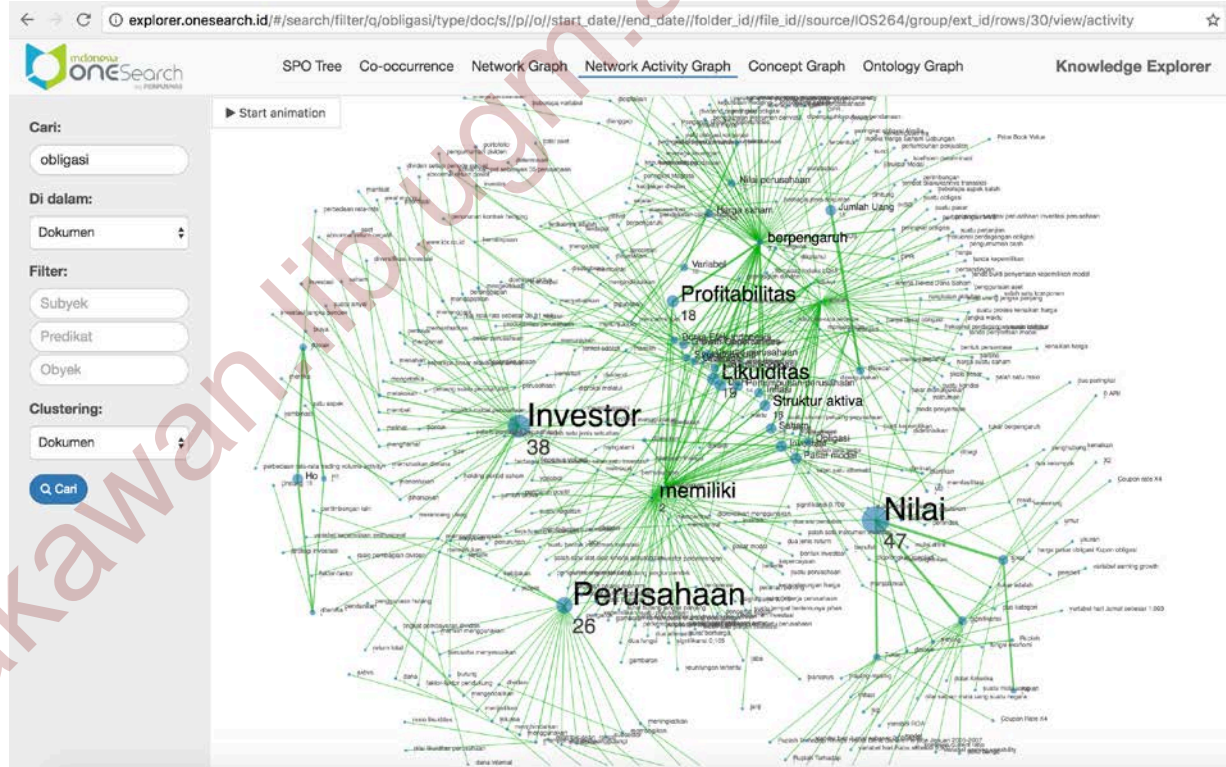
# Network Graph

- Klik menu 'Network Graph', dan akan ditampilkan relasi antara Subyek dan Object dalam triple yang berhasil diekstrak dari dokumen.
- Predikat tidak ditampilkan, dan ditunjukkan sebagai garis penghubung saja.
- Pada gambar di samping, kita lihat Network Graph dari dokumen yang mengandung kata kunci 'obligasi'.
- Tampak mana subyek atau obyek yang paling dominan dalam dokumen.



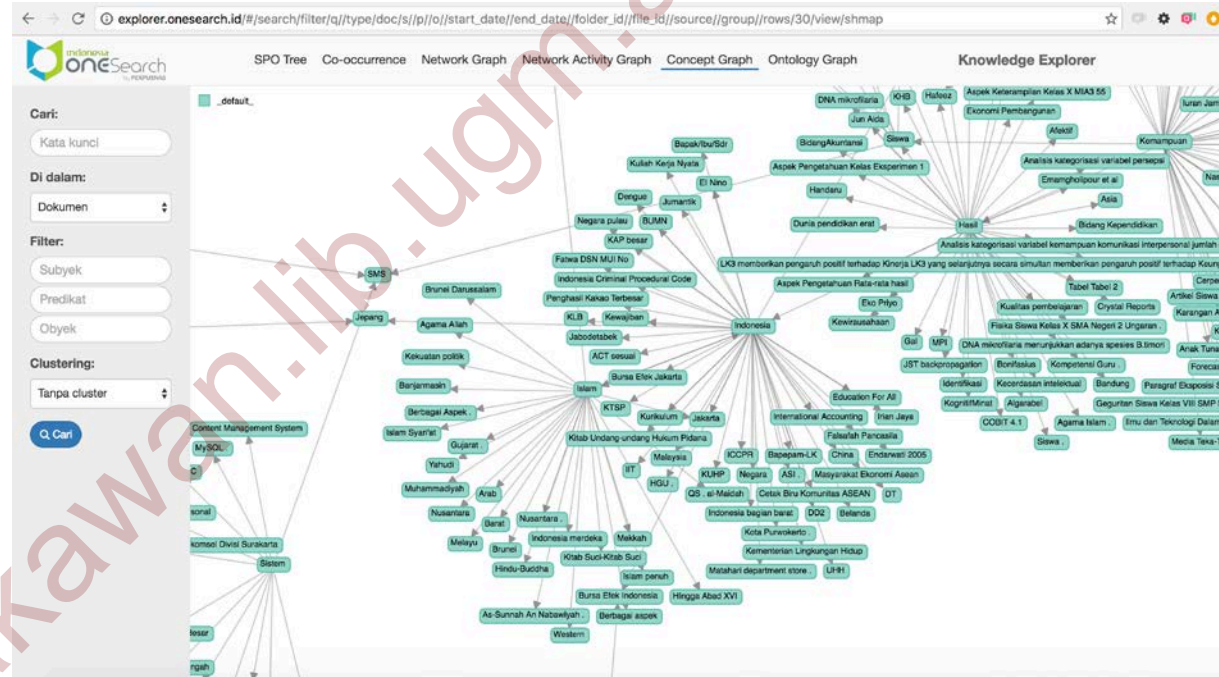
# Network Activity Graph

- Klik menu 'Network Activity Graph', maka akan ditampilkan relasi antara Subyek, Predikat, dan Obyek, yang masing-masing sebagai node.
- Contoh di samping adalah Activity graph dari triples yg ditemukan dalam dokumen yang mengandung kata kunci 'obligasi'.



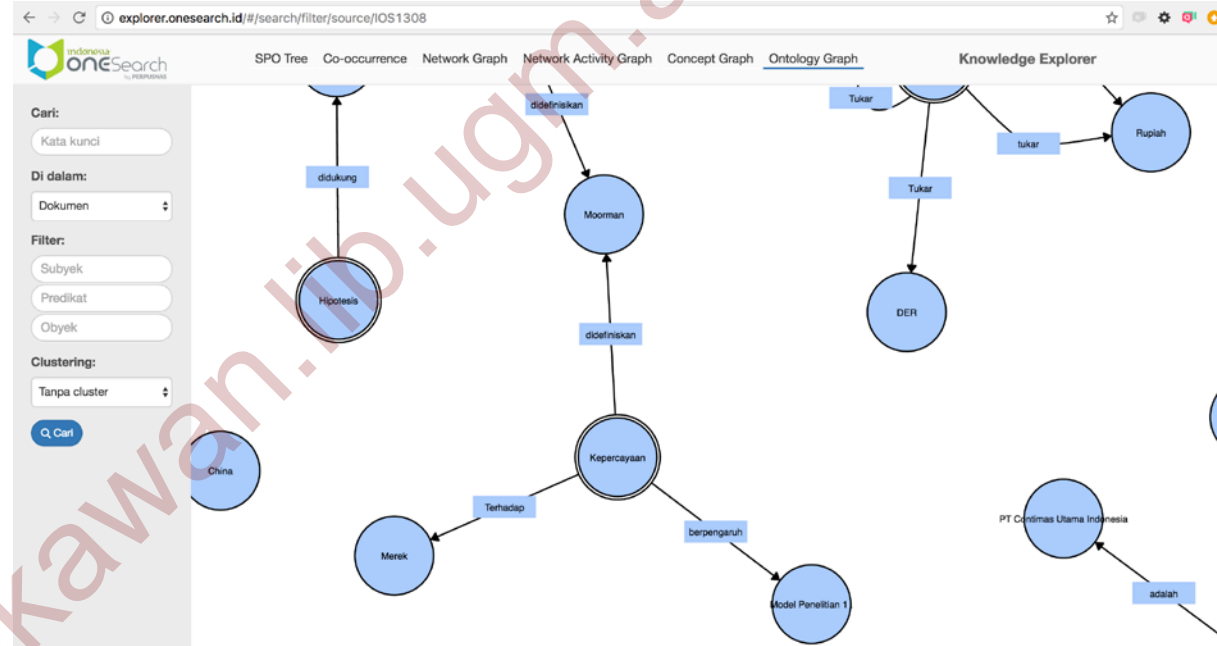
# Concept Graph

- Klik menu 'Concept Graph', akan ditampilkan relasi antara Subyek dan Obyek, melalui garis (predikat) yang memiliki arah.
- Predikat tidak ditampilkan labelnya di sini.
- Ini mirip dengan Network Graph, bedanya yang diperlihatkan adalah label atau nama yang lebih jelas dan relasi antara mereka.



# Ontology Graph

- Klik menu 'Ontology Graph', akan ditampilkan relasi antara Subyek dan Obyek, dengan predikat sebagai label dari garis yang menghubungkan.
- Di samping ini contoh beberapa graph yang ditemukan dalam sebuah repositori.

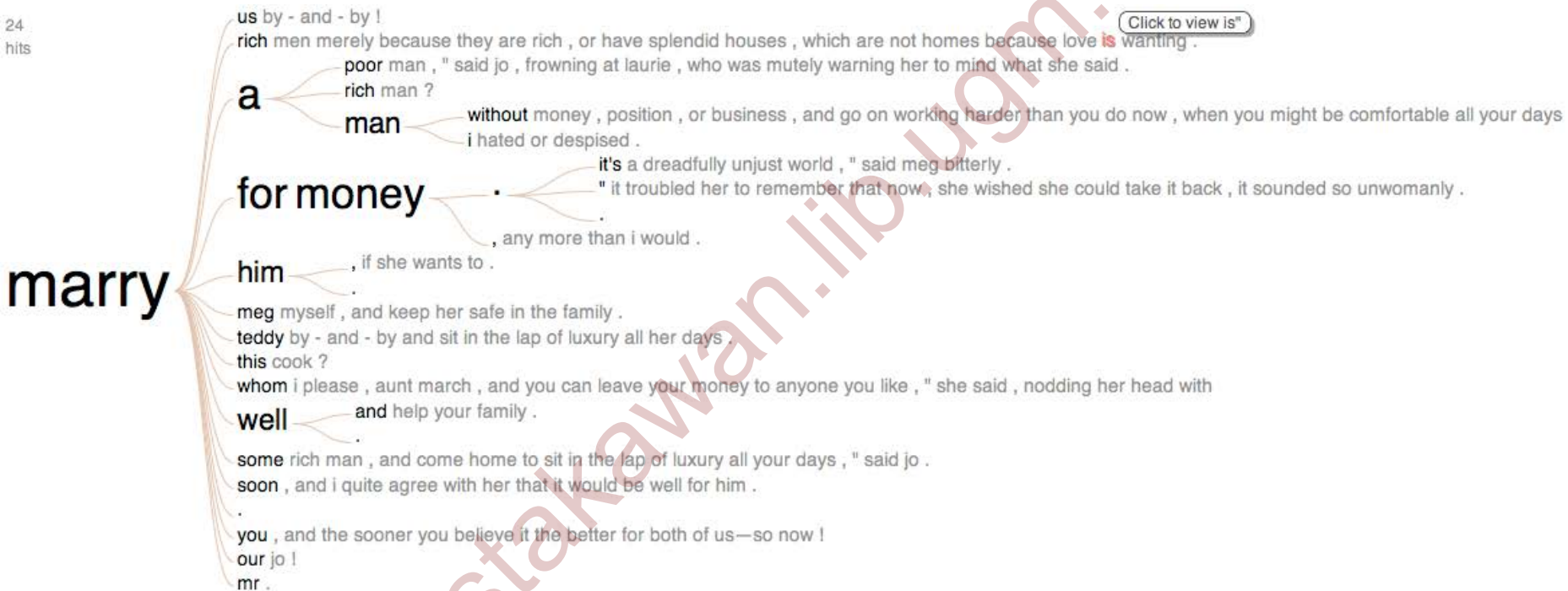


# Advanced Text Analysis

Experimental in IOS

# Word Tree

24  
hits

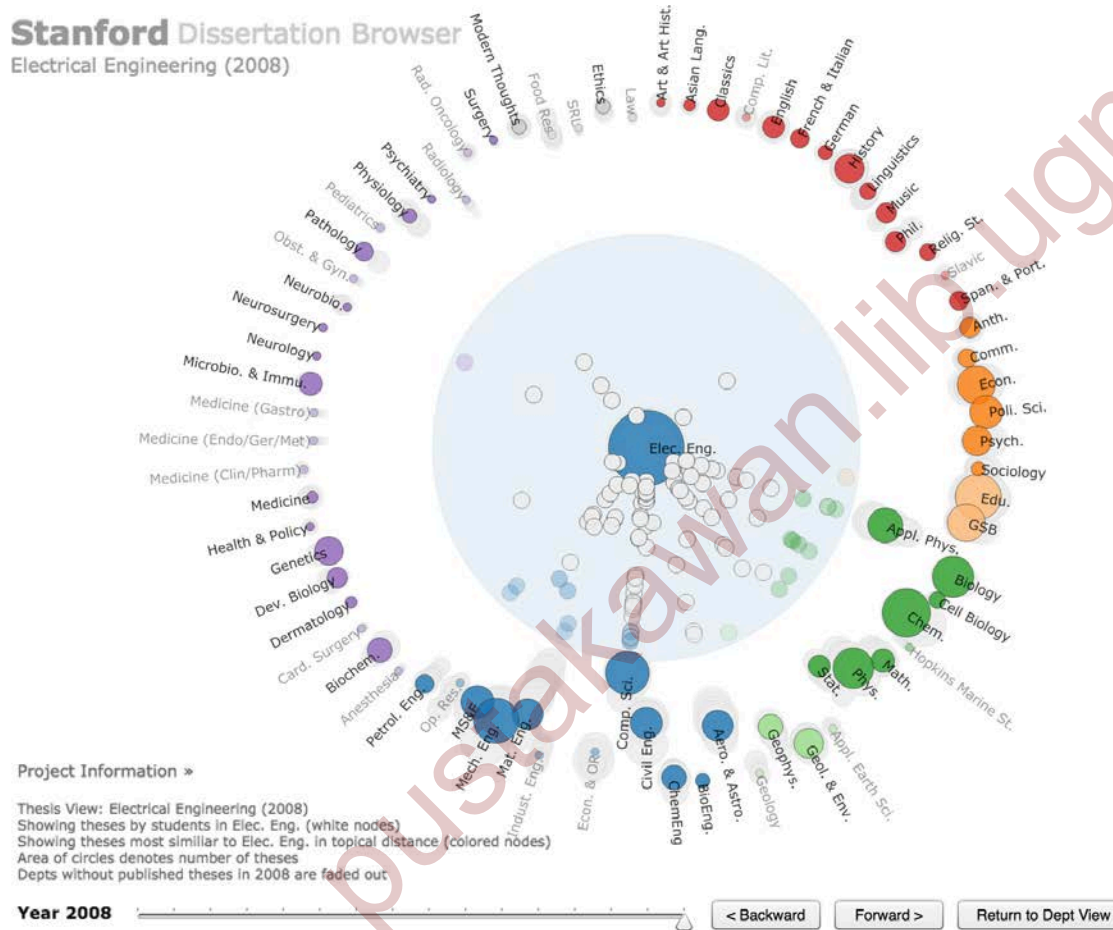




# Dissertation Browser

## Stanford Dissertation Browser

Electrical Engineering (2008)



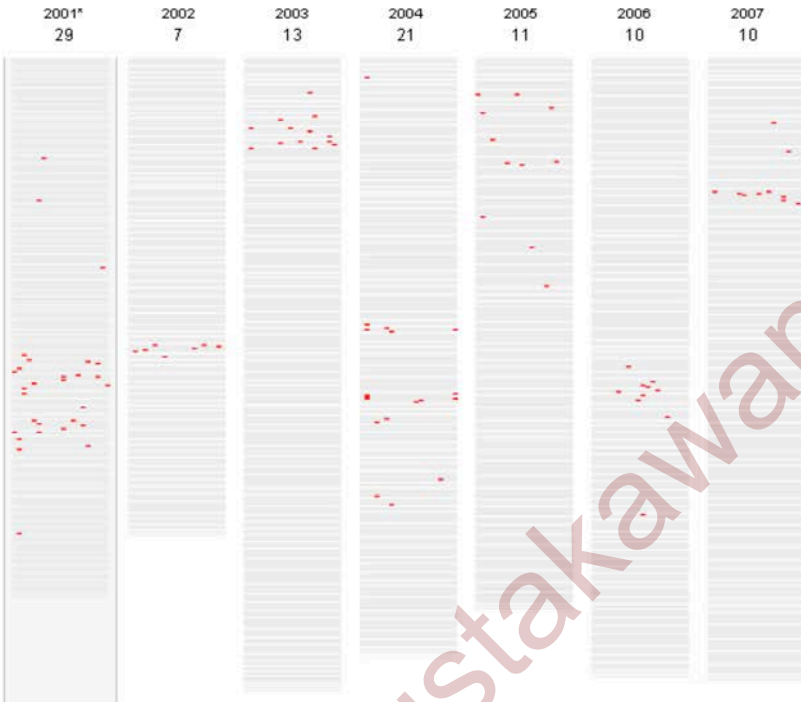
Stanford's PhD dissertation abstracts from 1993-2008 are presented through the lens of a text model that distills high-level similarity and word usage patterns in the data. You'll see each Stanford department as a circle, colored by school and sized by the number of PhD students graduating from that department.

When you click a department, it becomes the focus of the browser and every other department moves to show its relative similarity to the centered department.

# Phrase Analysis

 Search 

## Use of the phrase "Tax" in past State of the Union Addresses

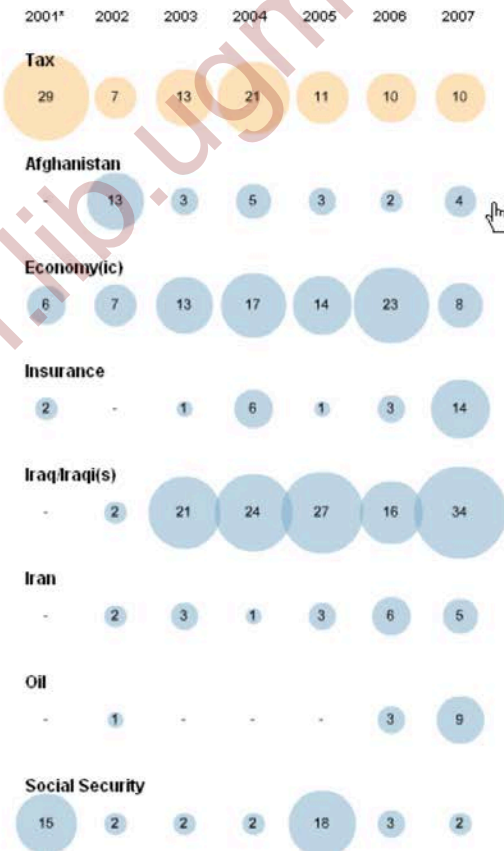


### The word in context

I believe in local control of schools. We should not, and we will not, run public schools from Washington, D.C. Yet when the federal government spends **TAX** dollars, we must insist on results. Children should be tested on basic reading and math skills every year between grades three and eight. Measuring is the only way to know whether all our children are learning. And I want to know, because I refuse to leave any child behind in America.

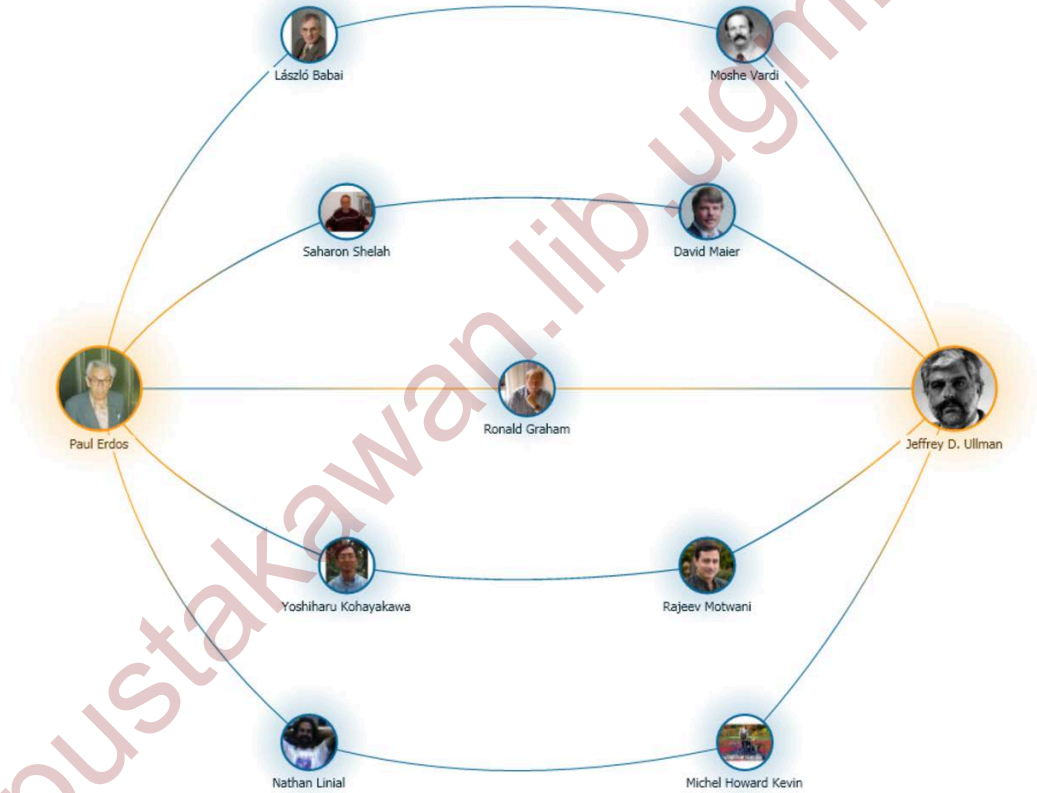
-- 2001 (Paragraph 14 of 73)

## Compared with other words



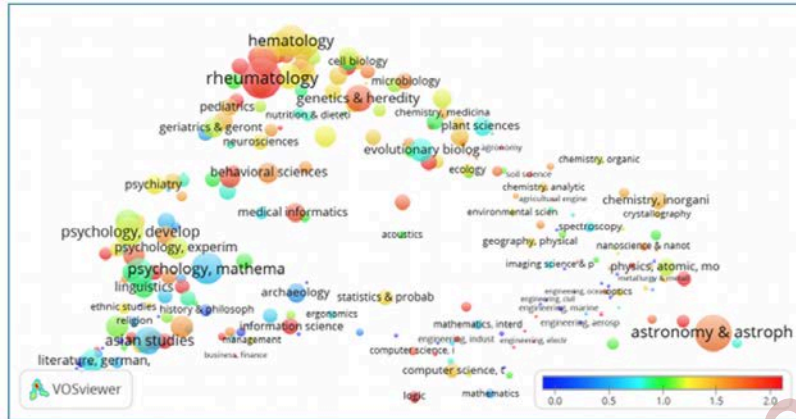
Depiction of word usages in the U.S. president's annual "State of the Union" speech across time. On the left, a visualization of the text is linked to the search box, and on the right, areas of circles show the relative frequency of word usage across time.

# Academic Collaboration Distance

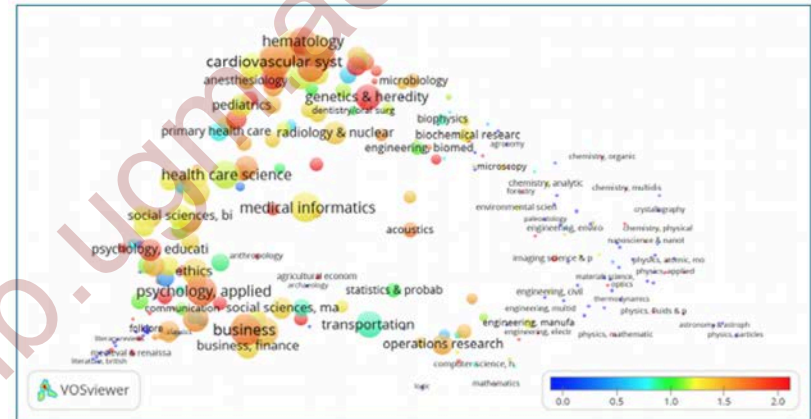


pustakawan.lib.ugm.ac.id

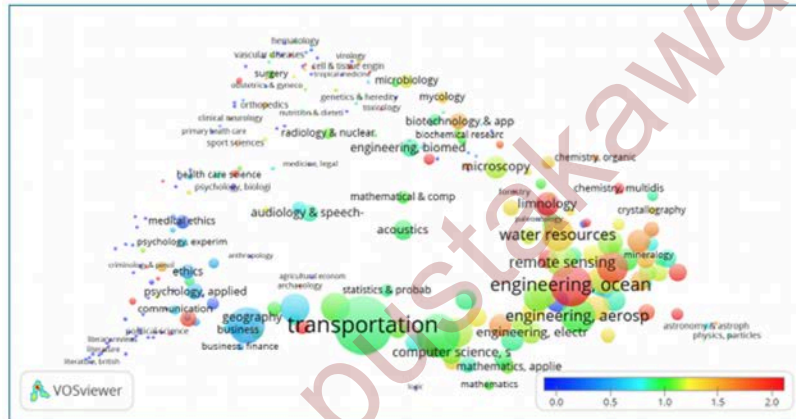
# University Profile Maps



Leiden University



Erasmus University Rotterdam



Delft University of Technology



# Peluang Baru bagi Pustakawan

pustakawan.lib.ugm.ac.id

# Peluang dan Manfaat

---

- Pustakawan memiliki tool analisis teks yaitu IOS Knowledge Explorer, tanpa harus investasi hardware/software.
- Pustakawan dapat menganalisis fullteks karya institusi masing-masing, membandingkan dengan institusi lain.
- Pustakawan bisa memberi masukan dan analisis kepada civitas dan institusi tentang peta pengetahuan bidang dan institusi masing-masing.
- Peran pustakawan makin penting dalam mengelola dan menganalisa repositori pengetahuan.

# Skill Set

---

- Analisis Teks
- Statistik

pustakawan.lib.ugm.ac.id

# Kesimpulan

---

- Open Access Indonesia – Discovery System (OAI-DS), dibutuhkan untuk mengelola seluruh repositori Open Access dan seluruh jenis data (tidak terbatas pada e-jurnal).
- Indonesia OneSearch (IOS) adalah sebuah Discovery System.
- Text Analysis dan Fact Extraction dalam IOS merupakan langkah awal membangun Knowledge Graph dari seluruh repositori Open Access di Indonesia.
- Pustakawan bisa memanfaatkan layanan Knowledge Explorer untuk menganalisa pengetahuan institusinya.
- Pustakawan bisa meningkatkan peran dan kontribusi berupa analisa pengetahuan bagi institusinya.



# Terimakasih

---

**Ismail Fahmi, PhD**

Email: [ismail.fahmi@gmail.com](mailto:ismail.fahmi@gmail.com)

Hp: 0812 8908 3894

pustakawan lib.ugm.ac.id