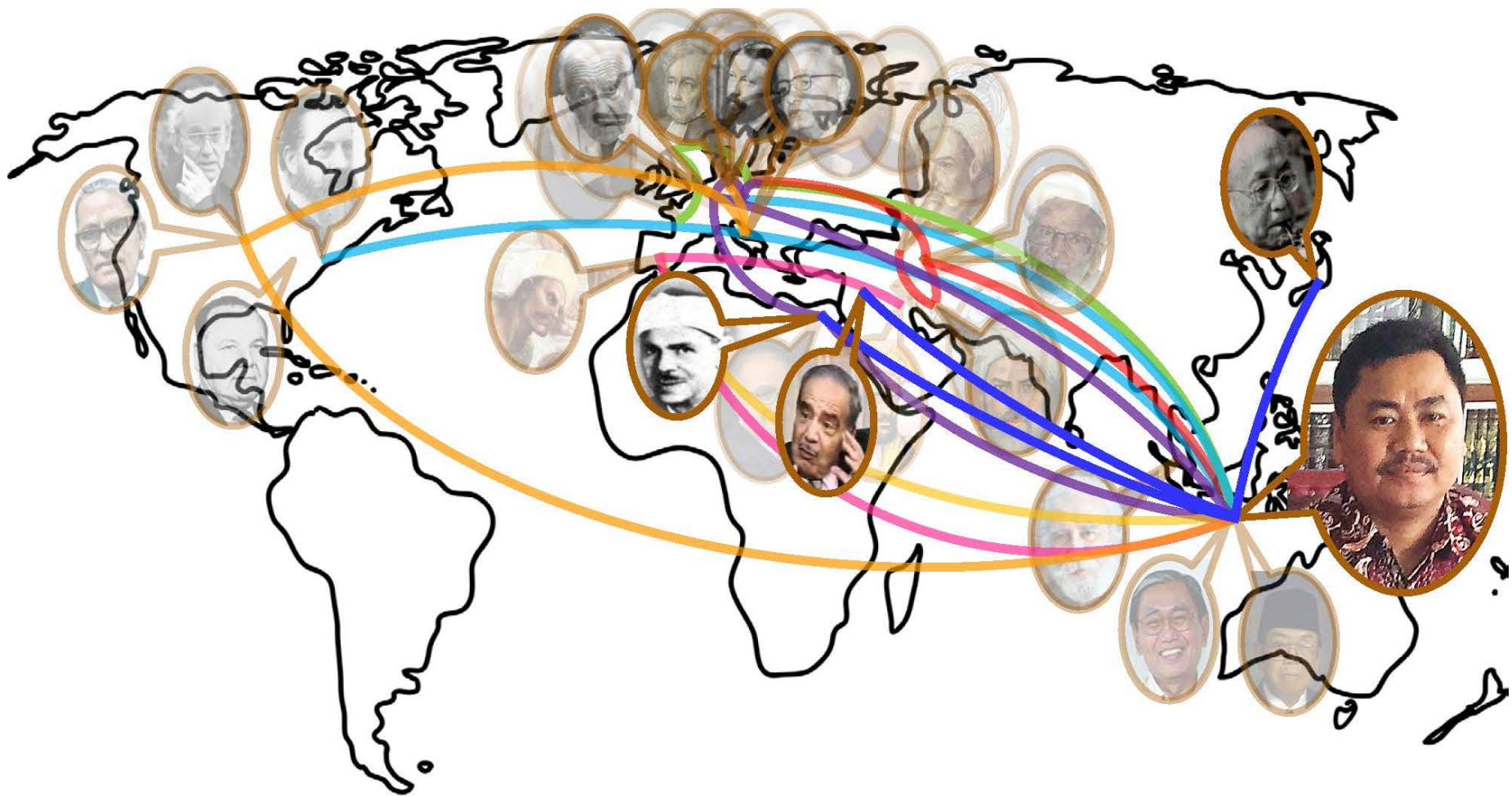


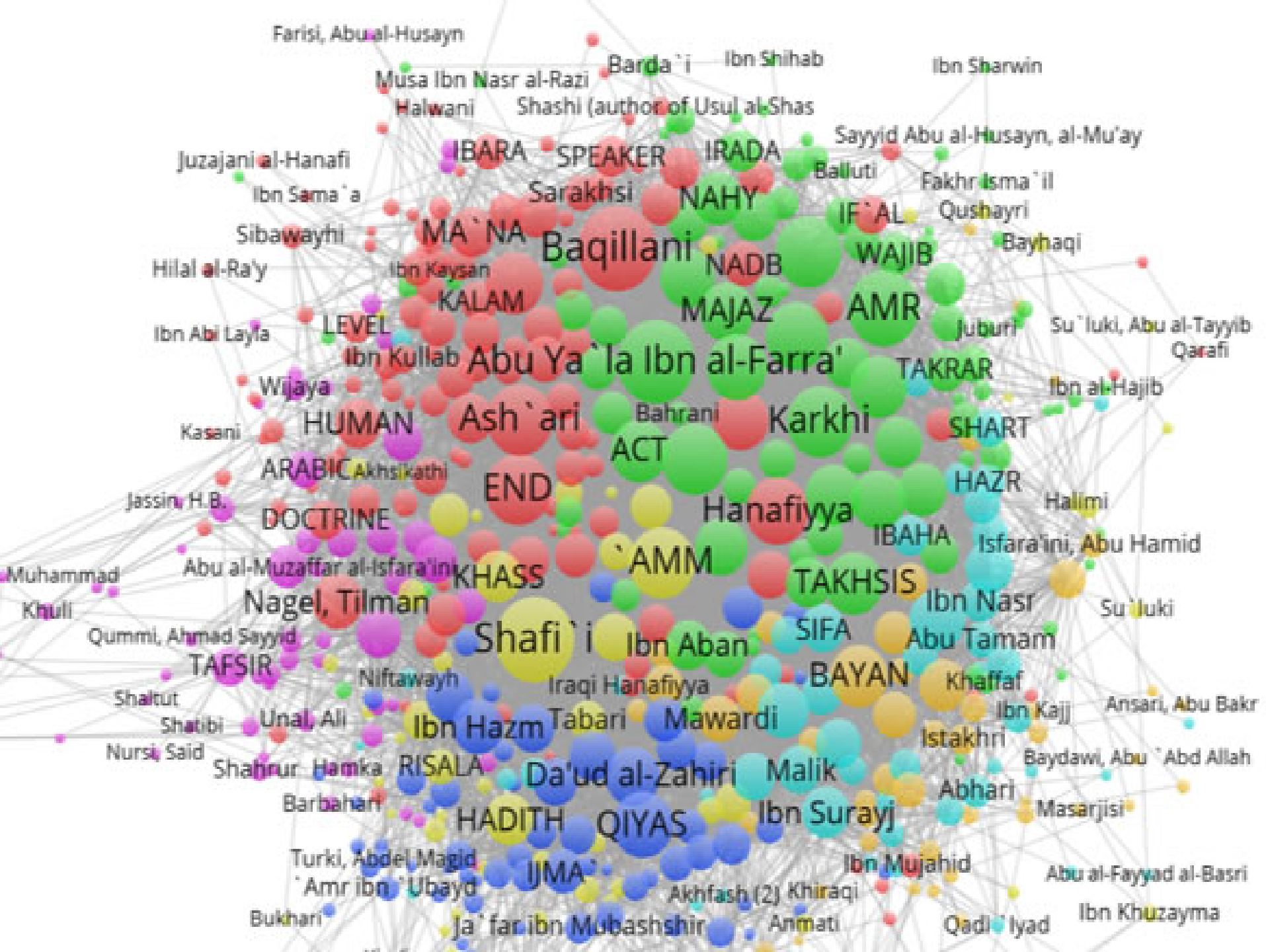
So Many Books, So Little Time: Using Algorithms to Map the Landscape of a Discourse

David Vishanoff, Associate Professor of Islamic Studies, University of Oklahoma
Dave King, Founder & Chief Executive Officer, Exaptive

Cairo, November 12, 2018

Bibliotheca Alexandrina and SIMAR conference “Big Data Analytics—Bridging the Gap between Theory and Practice”





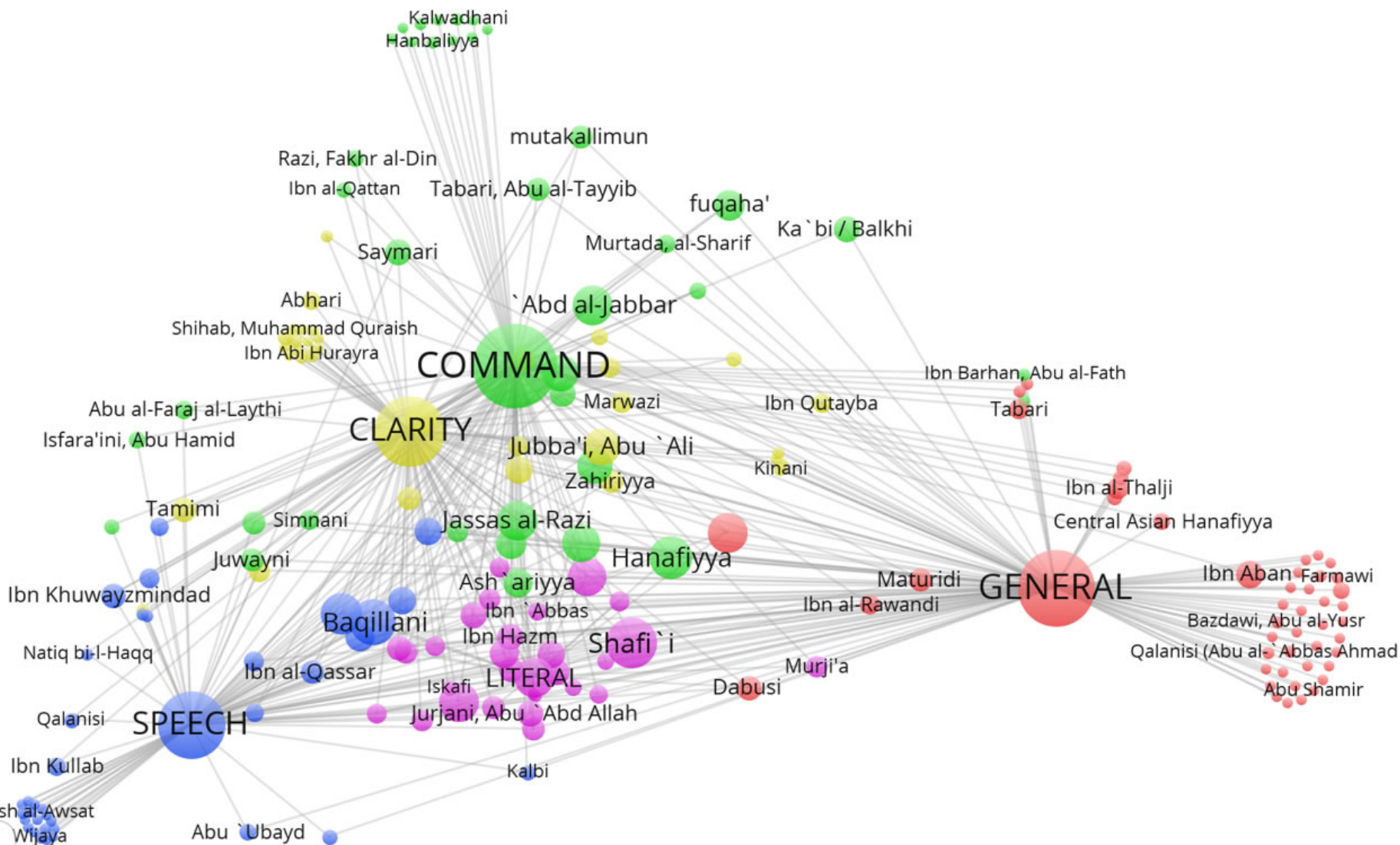
A word cloud of Arabic legal and linguistic terms, color-coded by region: purple (top), red (left), green (right), and blue (bottom). The terms are arranged in a circular pattern, with larger words indicating higher frequency or importance.

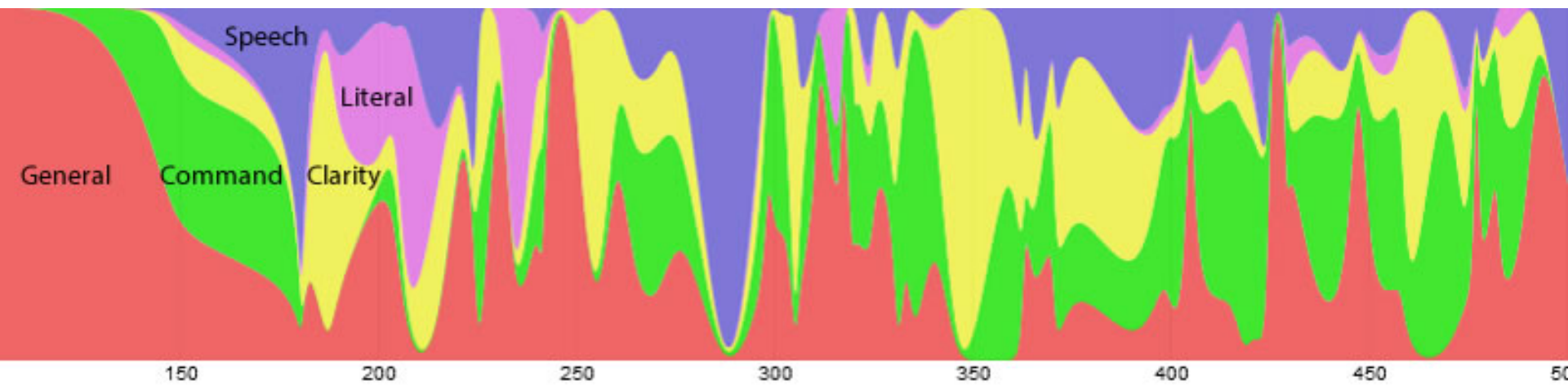
Purple Region (Top): dalil al khitab, ism, sifa, shart, mutlaq.

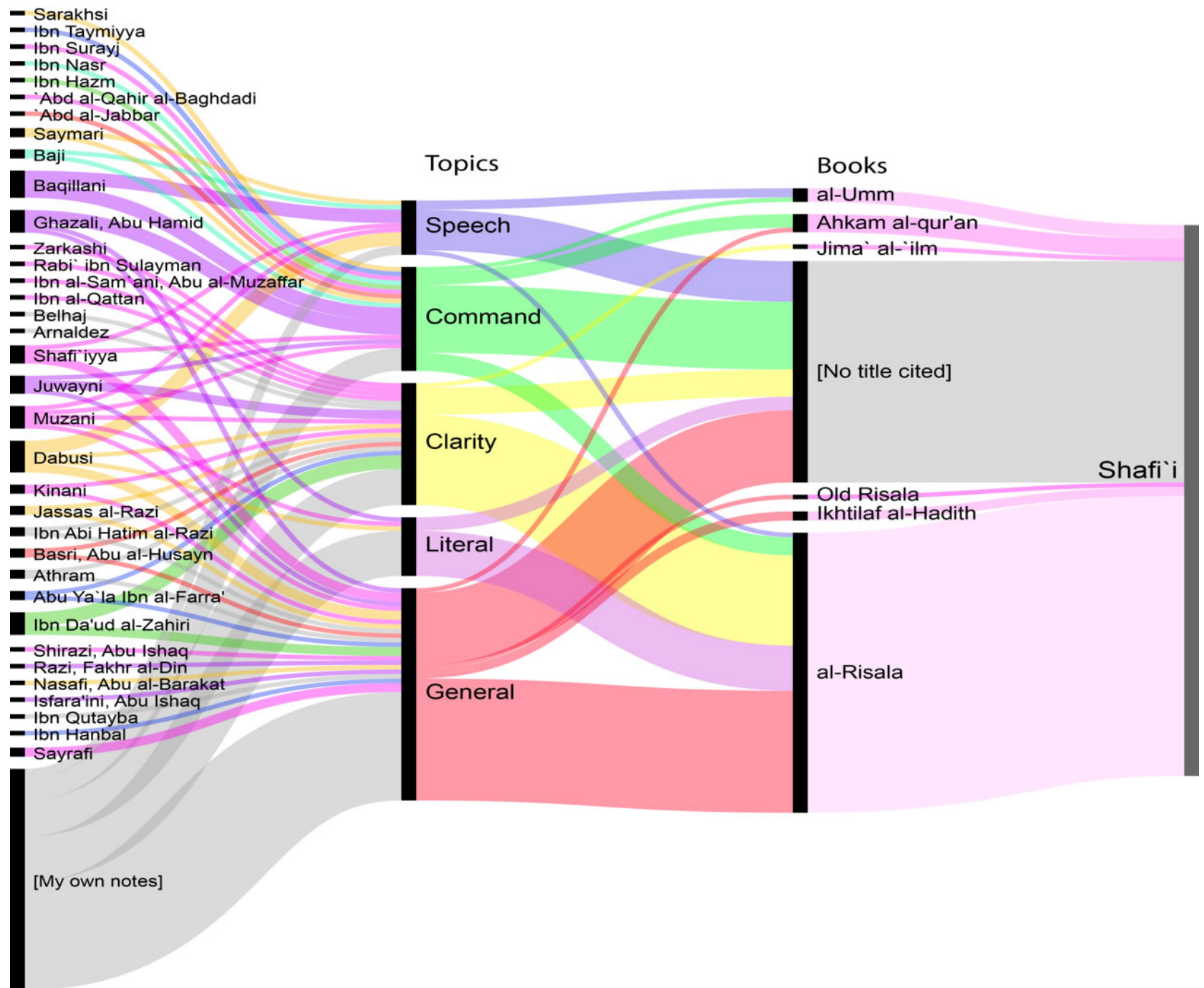
Red Region (Left): hujja negative implication, ijtihad, source, illa, qiyas, istidlal, hadith, tafsir, sunna, verse, woman, category, ambiguity, takhir al bayan, umum, waid, majaz, single utterance, haqiqa.

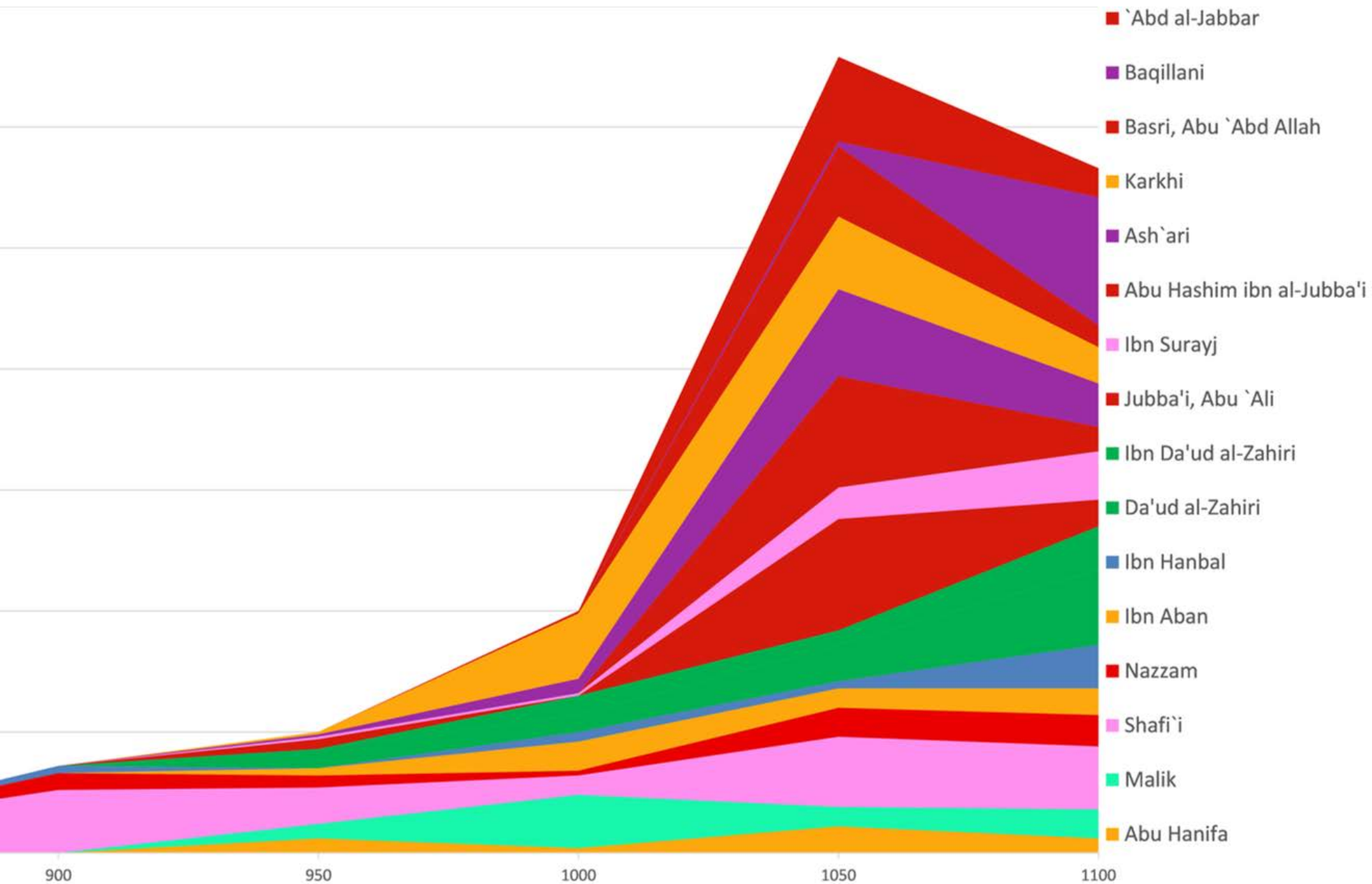
Green Region (Right): performance, ala al fawr, marra, takrar, wajib, mukallaf, act, amr, imperative, opposite, nadb, ifal, irada, qarina, prohibition.

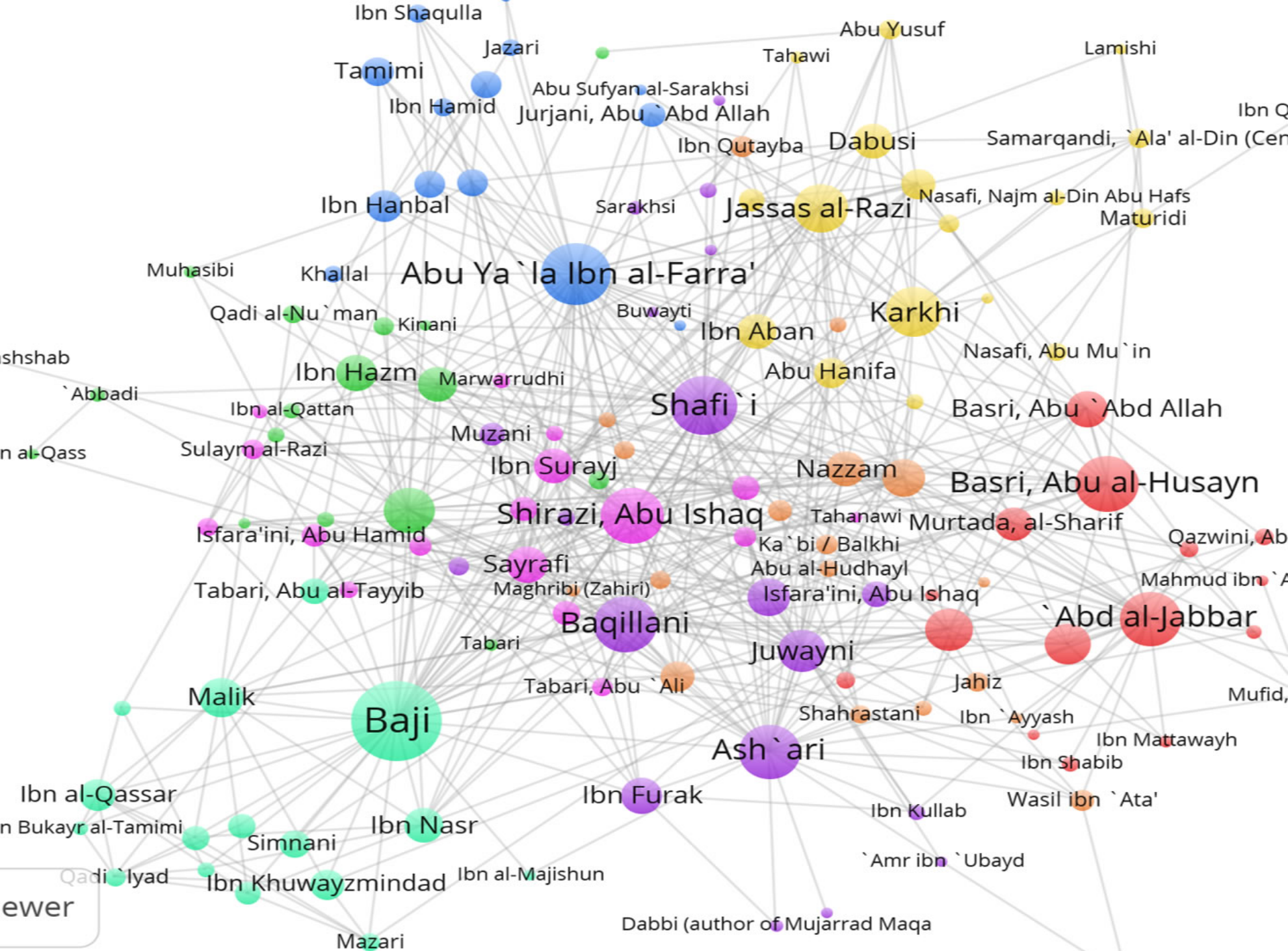
Blue Region (Bottom): speech, human speech, utterance, speaker, ibara, qasd, murad, human, number, sabab, beginning, end, debate, topic, doctrine, naskh, takhsis, amm, search, wahid, nash.











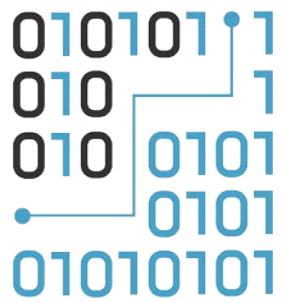
أخبرنا
السنة
شأن الشافعي
ذكر
الحجة
شيئا
نهى
أمره
أمر
حكم
عمر
أعلم
صلاة
وجه
معنى
وصفت
الكتاب
القياس
علم
دلالة
كتاب
الناس
شيء
أربعة
المعنى
فرض
سفيان
العلم
الصلاة
أولى
أحد
الرجل
يحتمل
أكثر
كتابه
القبلة
الآية
سمع
حرم
يحل
البقرة
القرآن
سعيد
خبر

0101011
0101
0100101
0101
01010101

Data



Algorithm



Data



Algorithm





Data



Algorithm



Human Collaboration





Data



Algorithm



Visualization



Human Collaboration



Insight

0101011
0101
0100101
0101
01010101

Data



Algorithm



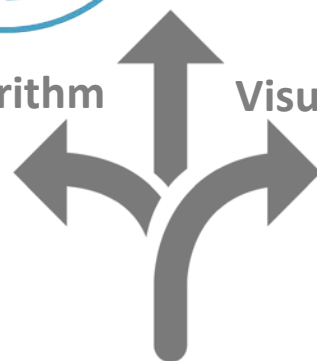
Visualization



Human Collaboration



Insight



Exaptation

0101011
0101
0100101
0101
01010101

Data



Algorithm



Visualization



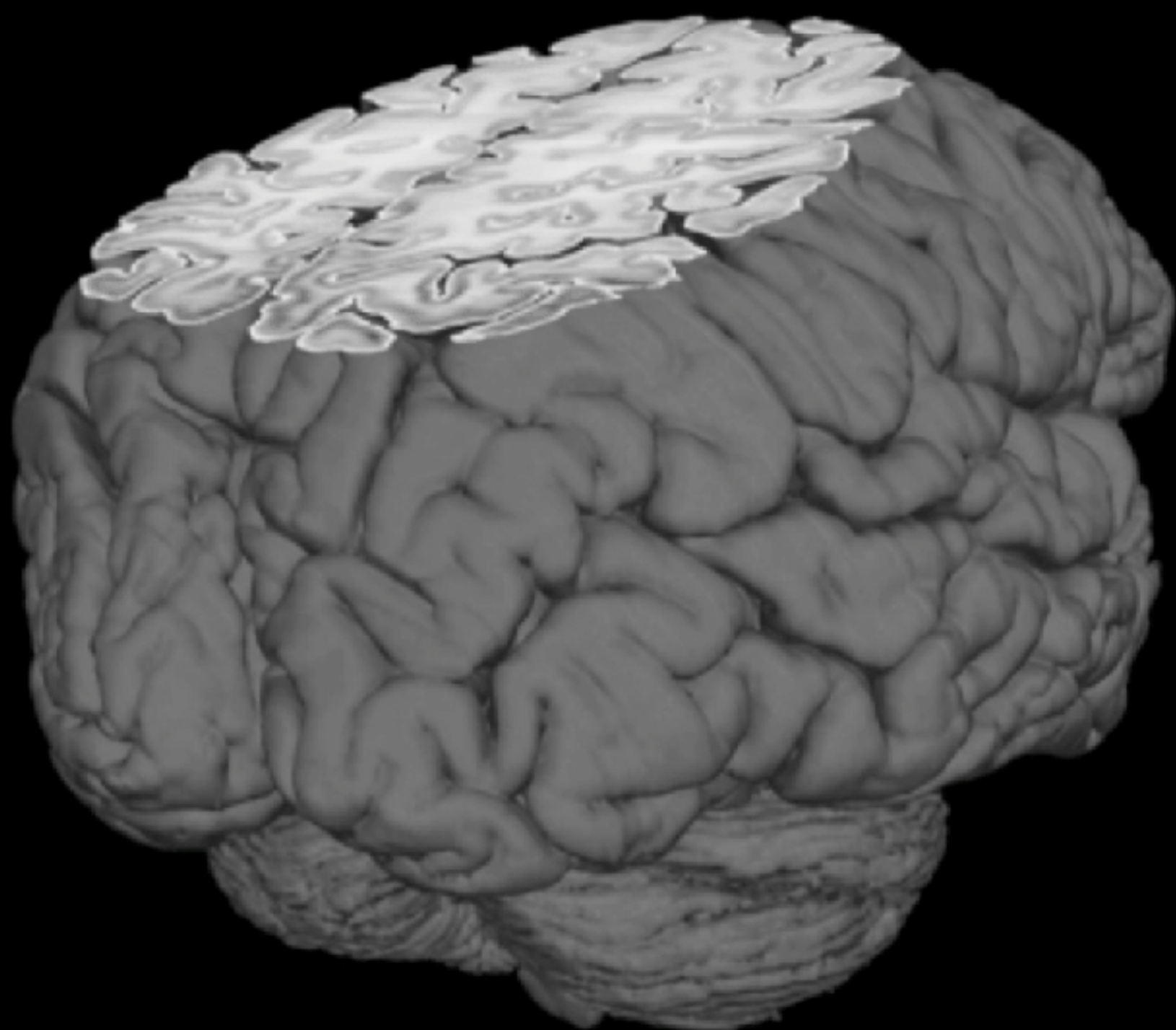
Human Collaboration

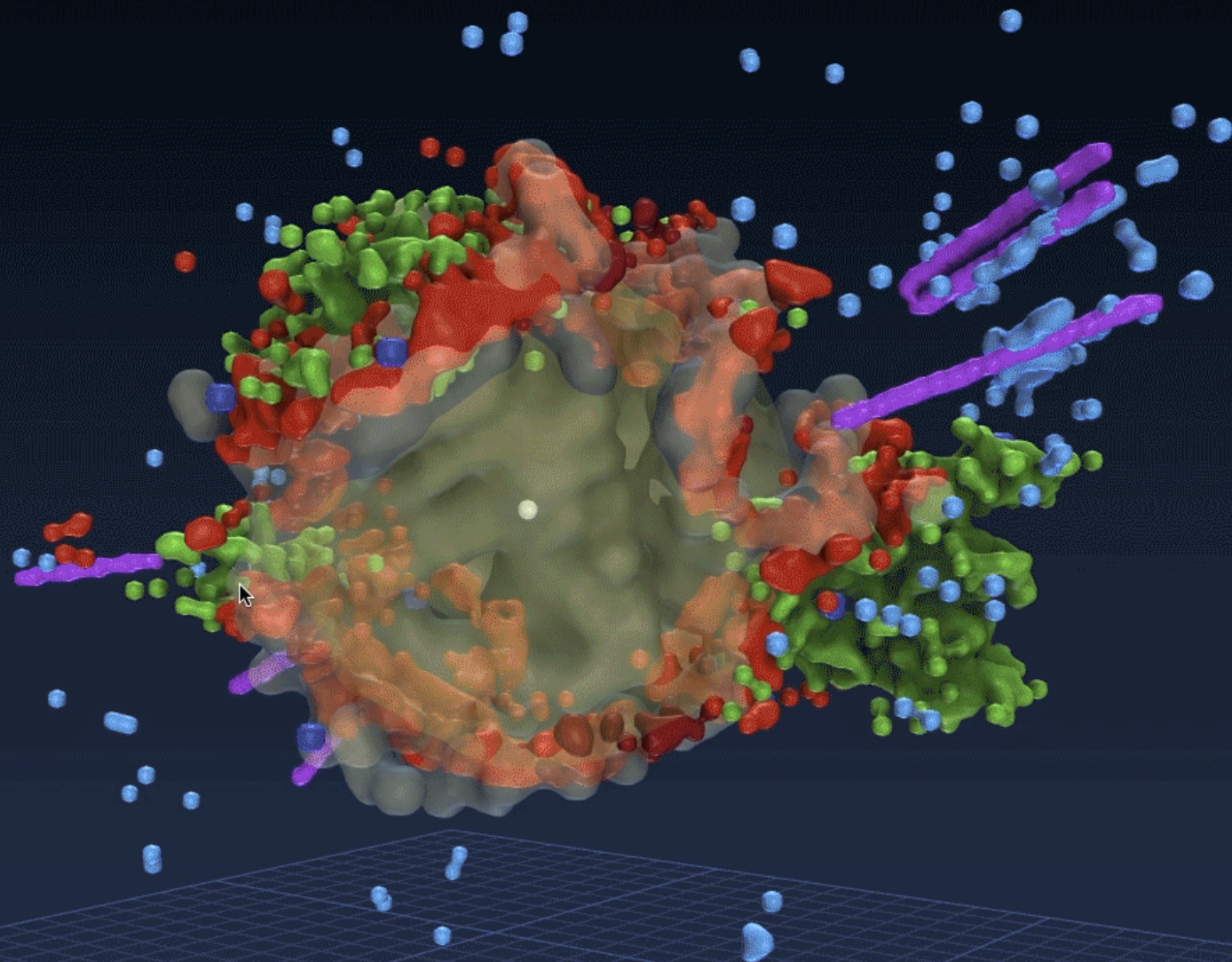


Insight

Exaptation

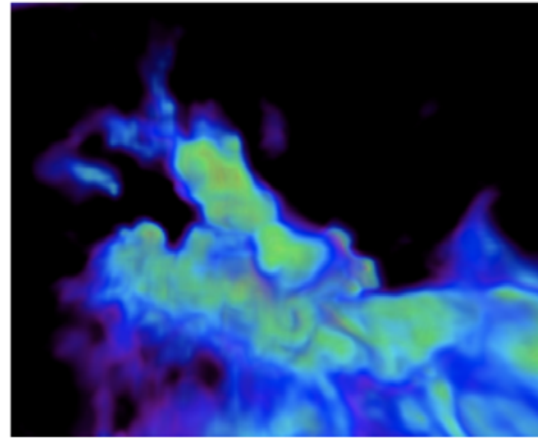
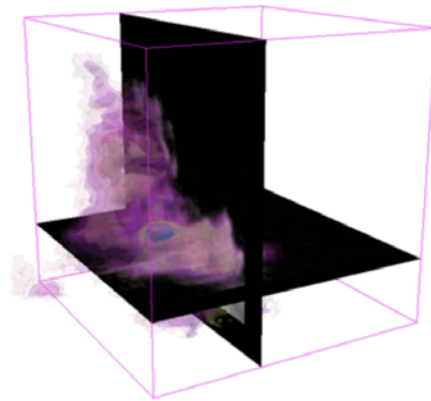






The Astronomical Medicine (AstroMed) Project

Douglas Alan (IIG), Michelle Borkin (IIG), Alyssa Goodman (IIG/CfA), Mike Halle (IIG/HMS), Jens Kauffmann (CfA/IIG)



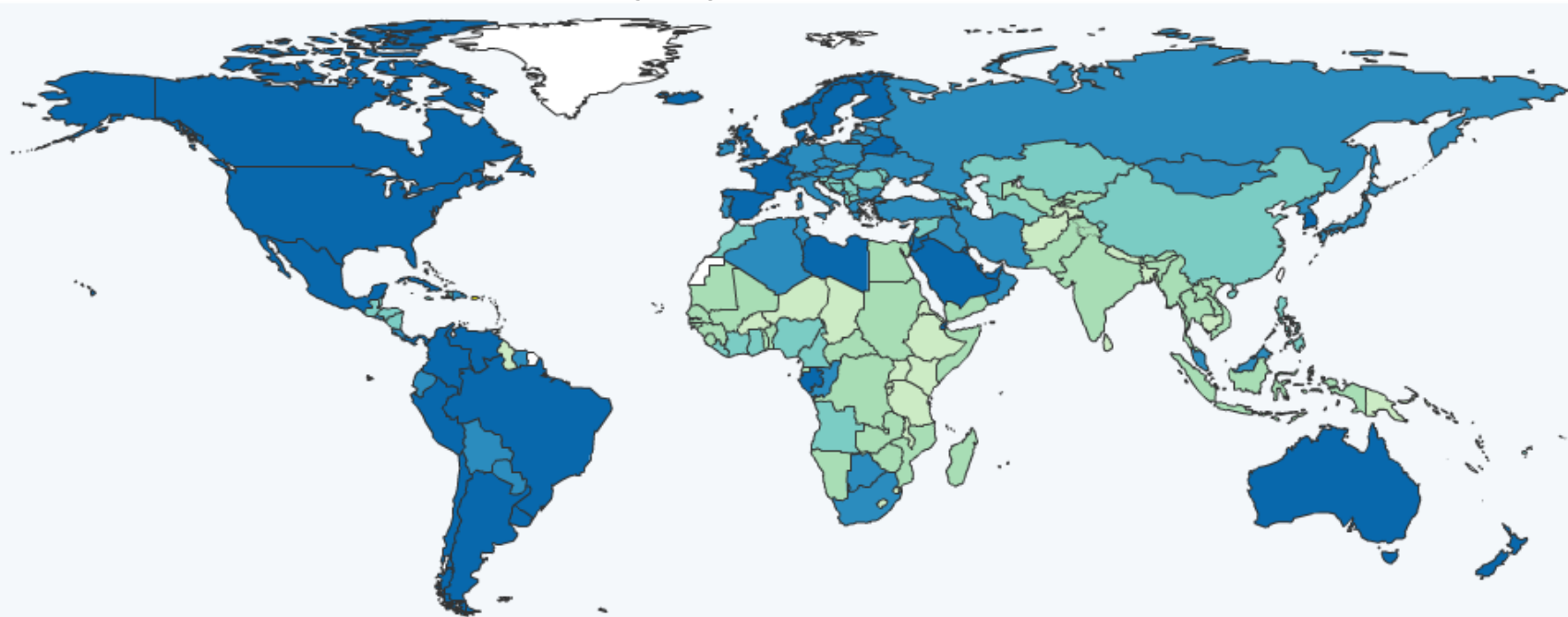
Views of the star forming region IC 348 as seen in the medical imaging programs 3D Slicer and OsiriX. Left: 3D Contour map of IC 348 with orthogonal slices as seen in 3D Slicer using ^{13}CO molecular line emission data from the COMPLETE Survey of Star Forming regions. Right: 3D Volume rendered model of IC 348 as seen in OsiriX using CO molecular line emission data from the COMPLETE Survey of Star Forming Regions.

We have a new project page! Please go to:
<http://astromed.iic.harvard.edu>

This page contains information about the medical imaging software used by the [AstroMed](#) project and the analysis of the COMPLETE data to date. The two programs currently being used are [3D Slicer](#) and [OsiriX](#) which can be downloaded for free. In the tables below is information for obtaining these programs, and a [sample set of data](#) with models. For more detailed information on the astronomical applications of 3D Slicer, see the publications listed below. If you have any questions, please let [us](#) know.

أخبرنا
السنة
شأن الشافعي
ذكر
الحجة
شيئا
نهى
أمره
أمر
حكم
عمر أعلم صلاة وجه
وصفت الكتاب
معنى
العلم
فرض سفيان
الصلوة
كتاب الناس
شيء أربعة
شأن
أولى
أحد
الرجل
يحتمل
أكثر
كتابه
القبلة
الآية
سمع حرم
سعيد
خبر
الظاهر

Population, urban (%)
(2011)



Example of tf-idf [\[edit \]](#)

Suppose we have term frequency tables for a collection consisting of only two documents, as listed on the right, then calculation of tf-idf for the term "this" in document 1 is performed as follows.

Tf, in its basic form, is just the frequency that we look up in appropriate table. In this case, it's one.

Idf is a bit more involved:

$$\text{idf}(\text{this}, D) = \log \frac{N}{|\{d \in D : t \in d\}|}$$

The numerator of the fraction is the number of documents, which is two. The number of documents in which "this" appears is also two, giving

$$\text{idf}(\text{this}, D) = \log \frac{2}{2} = 0$$

So tf-idf is zero for this term, and with the basic definition this is true of any term that occurs in all documents.

A slightly more interesting example arises from the word "example", which occurs three times but in only one document. For this document, tf-idf of "example" is:

$$\text{tf}(\text{example}, d_2) = 3$$

$$\text{idf}(\text{example}, D) = \log \frac{2}{1} \approx 0.3010$$

$$\text{tfidf}(\text{example}, d_2) = \text{tf}(\text{example}, d_2) \times \text{idf}(\text{example}, D) = 3 \times 0.3010 \approx 0.9030$$

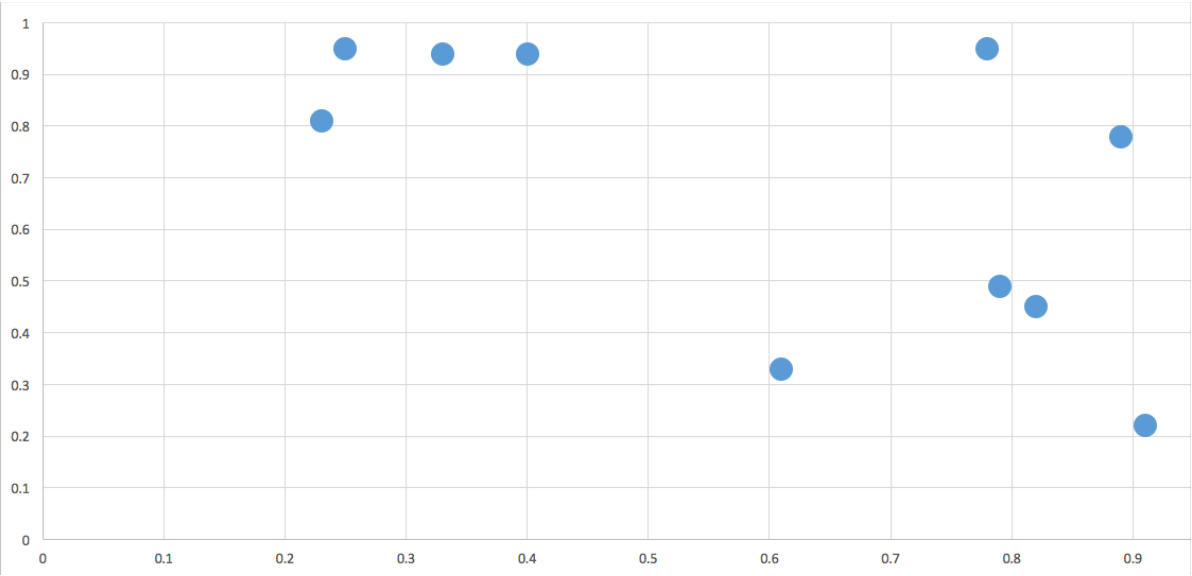
(using the [base 10 logarithm](#)).

Document 1		Document 2	
Term	Term Count	Term	Term Count
this	1	this	1
is	1	is	1
a	2	another	2
sample	1	example	3

	A	B	C	D	E	F	G	H	I	J	K
1		Doc1	Doc2	Doc3	Doc4	Doc5	Doc6	Doc7	Doc8	Doc9	Doc10
2	term1	0.78	0.95	0.54	0.73	0.58	0.51	0.36	0.52	0.92	0.65
3	term2	0.33	0.94	0.65	0.03	0.55	0.35	0.74	0.17	0.04	0.05
4	term3	0.89	0.78	0.35	0.58	0.98	0.56	0.84	0.65	0.25	0.03
5	term4	0.91	0.22	0.18	0.95	0.02	0.24	0.83	0.58	0.99	0.9
6	term5	0.25	0.95	0.03	0.98	0.9	0.19	0.91	0.04	0.23	0.96
7	term6	0.23	0.81	0.64	0.8	0.55	0.06	0.29	0.61	0.74	0.52
8	term7	0.79	0.49	0.33	0.8	0.5	0.23	0.32	0.79	0.03	0.34
9	term8	0.61	0.33	0.56	0.18	0.6	0.26	0.23	0.69	0.55	0.15
10	term9	0.4	0.94	0.57	0.98	0.24	0.54	0.33	0.89	0.34	0.89
11	term10	0.82	0.45	0.76	0.24	0.48	0.85	0.59	0.6	0.51	0.22

	A	B	C	D	E	F	G	H	I	J	K
1		Doc1	Doc2	Doc3	Doc4	Doc5	Doc6	Doc7	Doc8	Doc9	Doc10
2	term1	0.78	0.95	0.54	0.73	0.58	0.51	0.36	0.52	0.92	0.65
3	term2	0.33	0.94	0.65	0.03	0.55	0.35	0.74	0.17	0.04	0.05

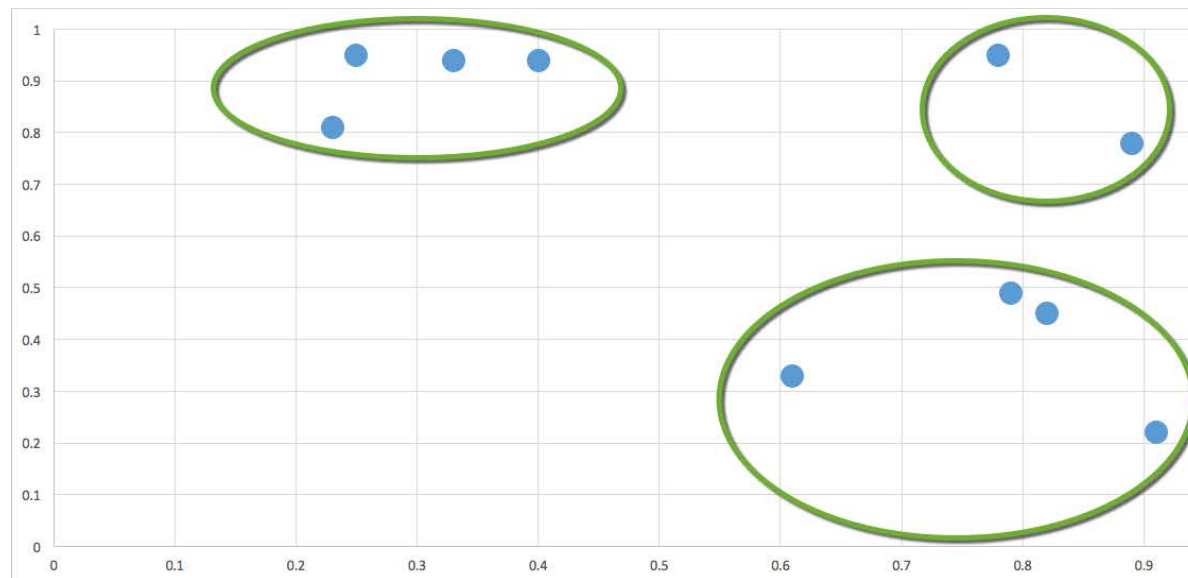
term1



term2

	A	B	C	D	E	F	G	H	I	J	K
1		Doc1	Doc2	Doc3	Doc4	Doc5	Doc6	Doc7	Doc8	Doc9	Doc10
2	term1	0.78	0.95	0.54	0.73	0.58	0.51	0.36	0.52	0.92	0.65
3	term2	0.33	0.94	0.65	0.03	0.55	0.35	0.74	0.17	0.04	0.05

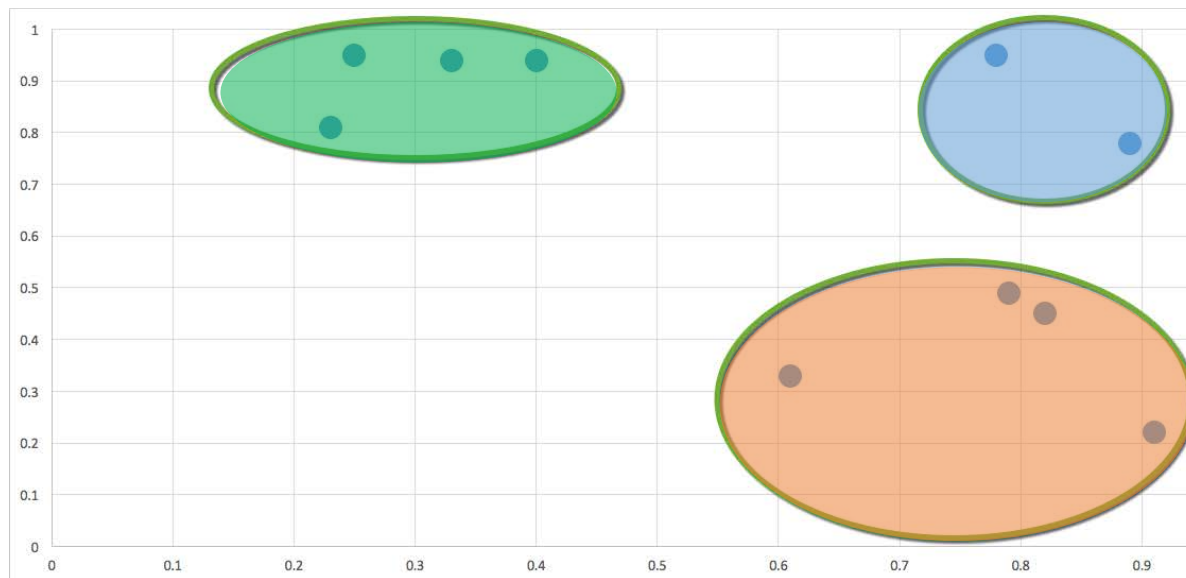
term1



term2

	A	B	C	D	E	F	G	H	I	J	K
1		Doc1	Doc2	Doc3	Doc4	Doc5	Doc6	Doc7	Doc8	Doc9	Doc10
2	term1	0.78	0.95	0.54	0.73	0.58	0.51	0.36	0.52	0.92	0.65
3	term2	0.33	0.94	0.65	0.03	0.55	0.35	0.74	0.17	0.04	0.05

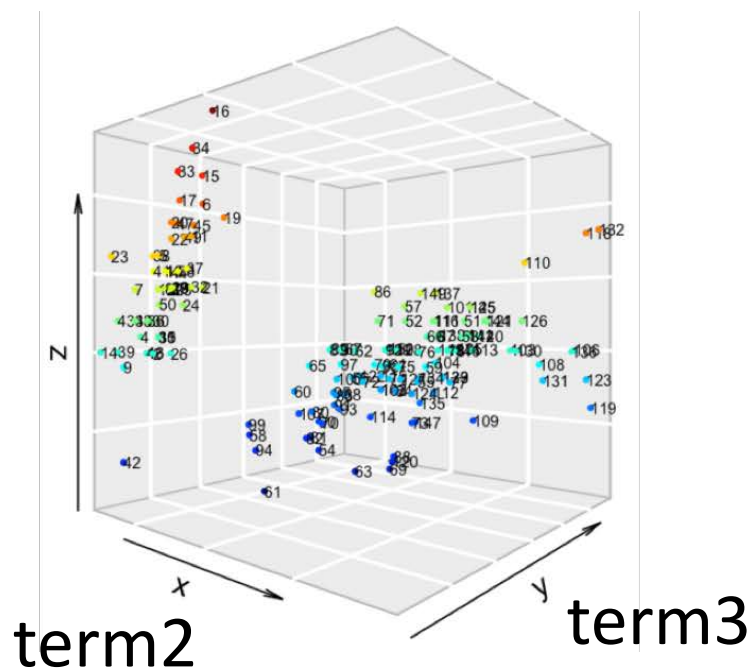
term1



term2

	A	B	C	D	E	F	G	H	I	J	K
1		Doc1	Doc2	Doc3	Doc4	Doc5	Doc6	Doc7	Doc8	Doc9	Doc10
2	term1	0.78	0.95	0.54	0.73	0.58	0.51	0.36	0.52	0.92	0.65
3	term2	0.33	0.94	0.65	0.03	0.55	0.35	0.74	0.17	0.04	0.05
4	term3	0.89	0.78	0.35	0.58	0.98	0.56	0.84	0.65	0.25	0.03

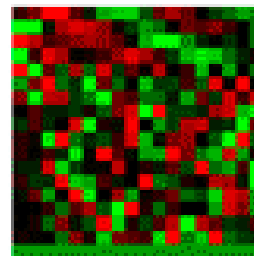
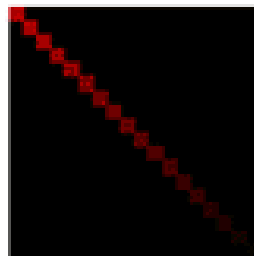
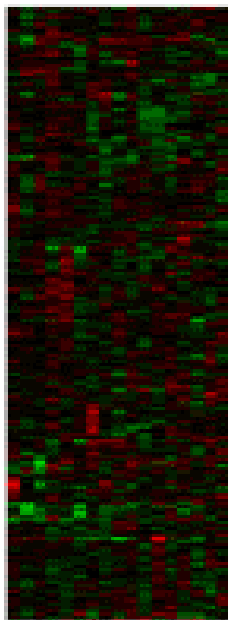
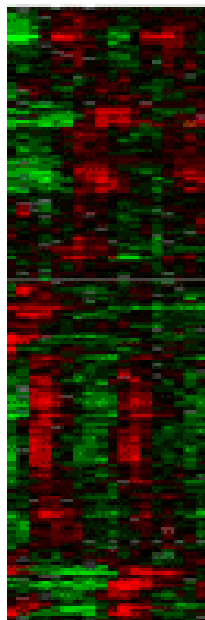
term1



	A	B	C	D	E	F	G	H	I	J	K
1		Doc1	Doc2	Doc3	Doc4	Doc5	Doc6	Doc7	Doc8	Doc9	Doc10
2	term1	0.78	0.95	0.54	0.73	0.58	0.51	0.36	0.52	0.92	0.65
3	term2	0.33	0.94	0.65	0.03	0.55	0.35	0.74	0.17	0.04	0.05
4	term3	0.89	0.78	0.35	0.58	0.98	0.56	0.84	0.65	0.25	0.03
5	term4	0.91	0.22	0.18	0.95	0.02	0.24	0.83	0.58	0.99	0.9
6	term5	0.25	0.95	0.03	0.98	0.9	0.19	0.91	0.04	0.23	0.96
7	term6	0.23	0.81	0.64	0.8	0.55	0.06	0.29	0.61	0.74	0.52
8	term7	0.79	0.49	0.33	0.8	0.5	0.23	0.32	0.79	0.03	0.34
9	term8	0.61	0.33	0.56	0.18	0.6	0.26	0.23	0.69	0.55	0.15
10	term9	0.4	0.94	0.57	0.98	0.24	0.54	0.33	0.89	0.34	0.89
11	term10	0.82	0.45	0.76	0.24	0.48	0.85	0.59	0.6	0.51	0.22

	A	B	C	D	E	F	G	H	I	J	K
1		Doc1	Doc2	Doc3	Doc4	Doc5	Doc6	Doc7	Doc8	Doc9	Doc10
2	term1	0.78	0.95	0.54	0.73	0.58	0.51	0.36	0.52	0.92	0.65
3	term2	0.33	0.94	0.65	0.03	0.55	0.35	0.74	0.17	0.04	0.05
4	term3	0.89	0.78	0.35	0.58	0.98	0.56	0.84	0.65	0.25	0.03
5	term4	0.91	0.22	0.18	0.95	0.02	0.24	0.83	0.58	0.99	0.9
6	term5	0.25	0.95	0.03	0.98	0.9	0.19	0.91	0.04	0.23	0.96
7	term6	0.23	0.81	0.64	0.8	0.55	0.06	0.29	0.61	0.74	0.52
8	term7	0.79	0.49	0.33	0.8	0.5	0.23	0.32	0.79	0.03	0.34
9	term8	0.61	0.33	0.56	0.18	0.6	0.26	0.23	0.69	0.55	0.15
10	term9	0.4	0.94	0.57	0.98	0.24	0.54	0.33	0.89	0.34	0.89
11	term10	0.82	0.45	0.76	0.24	0.48	0.85	0.59	0.6	0.51	0.22

$$A = U \cdot W \cdot V^T$$







Showing 100 out of 100 (35,384 available)

Climbing higher

by Williams, Montel.

In 1999, after almost twenty years of mysterious symptoms that he had tried to ignore, Montel Williams, a decorated former naval intelligence officer and Emmy Award-winning talk show host, was finally diagnosed with multiple sclerosis.

The top of the world : climbing Mount Everest

by Jenkins, Steve, 1952-

Describes the conditions and terrain of Mount Everest, attempts that have been made to scale this peak, and general information about the equipment and techniques of mountain climbing.

Climbing the stairs

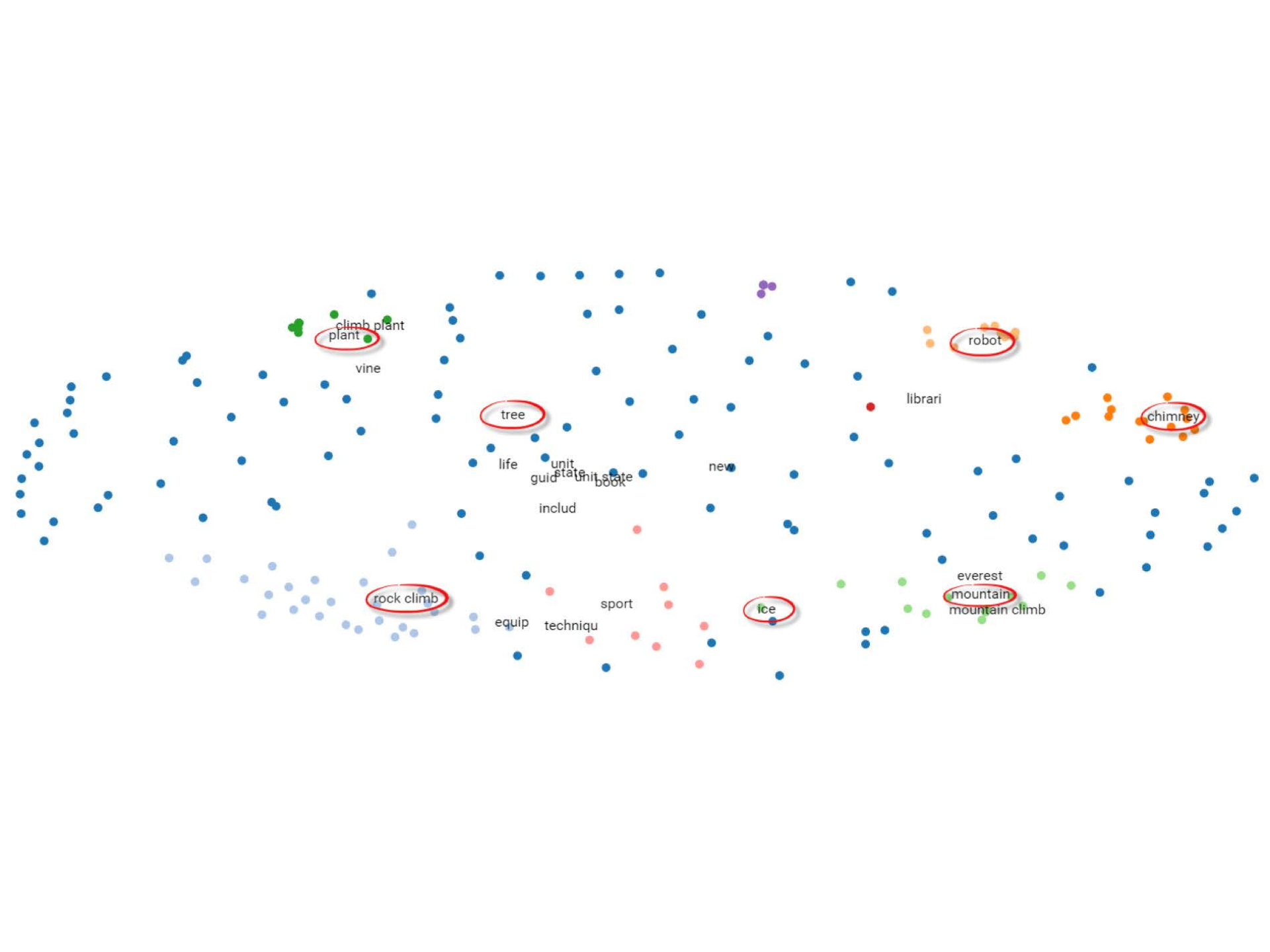
by Venkatraman, Padma.

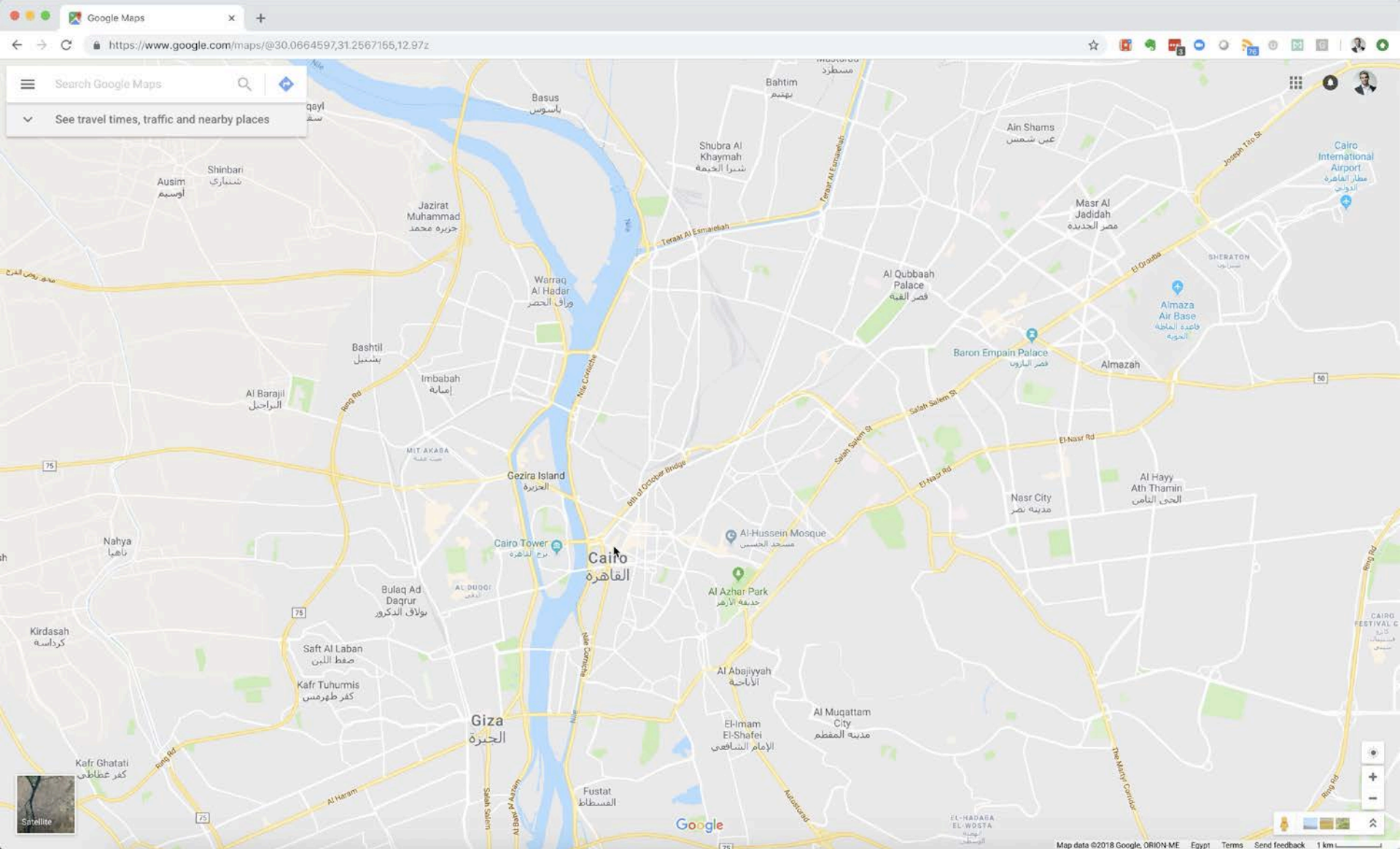
In India, in 1941, when her father becomes brain-damaged in a non-violent protest march, fifteen-year-old Vidya and her family are forced to move in with her father's extended family and become accustomed to a totally different way of life.

The case of the climbing cat

by Rylant, Cynthia.

The High Rise Private Eyes, animal detectives, try to find the cat who stole their neighbor's binoculars.





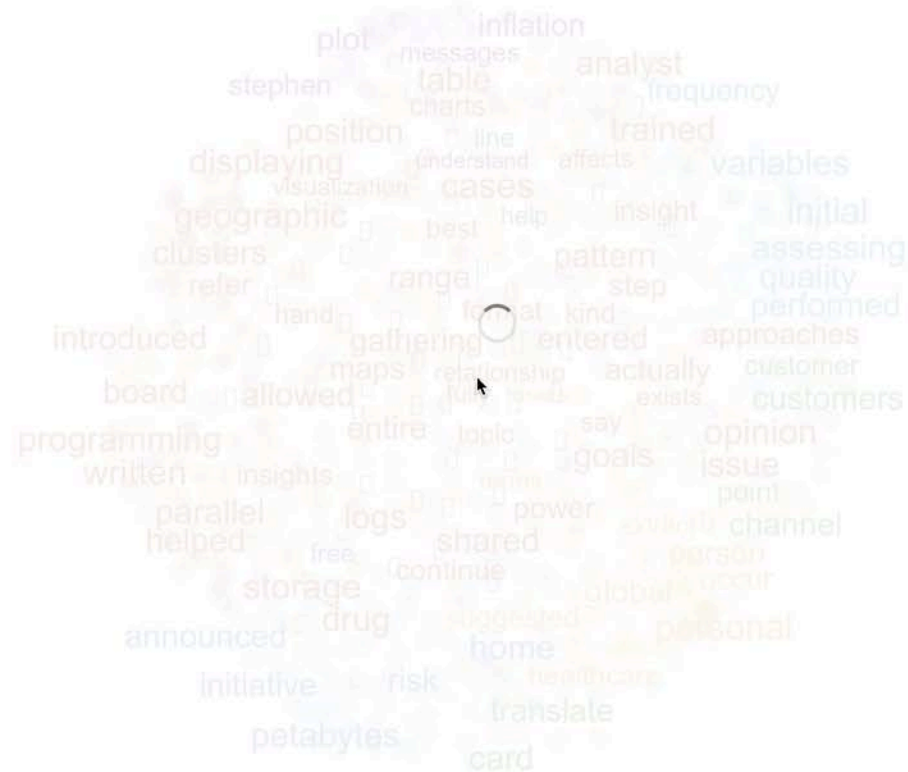
Select a Corpus:

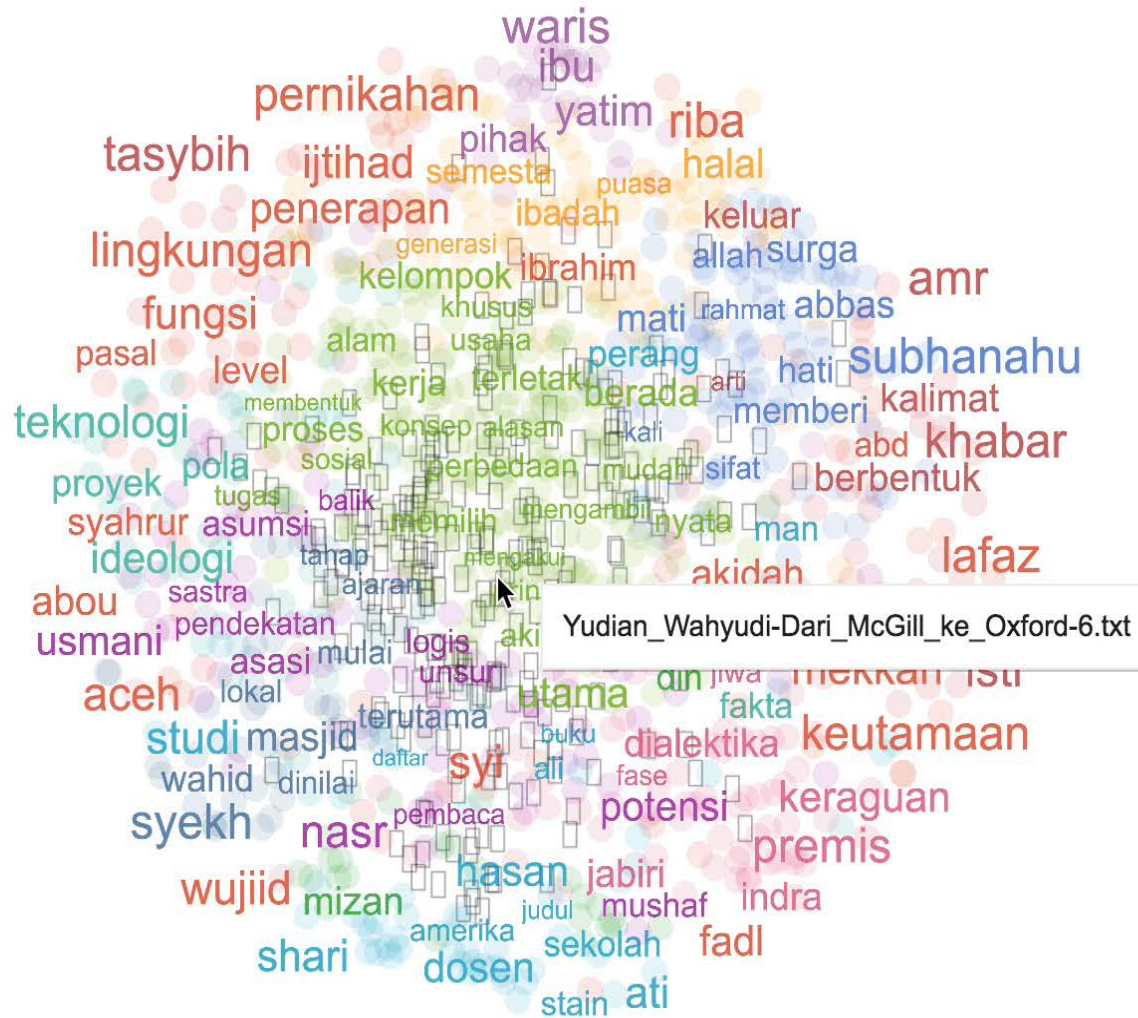
Wikipedia Big Data

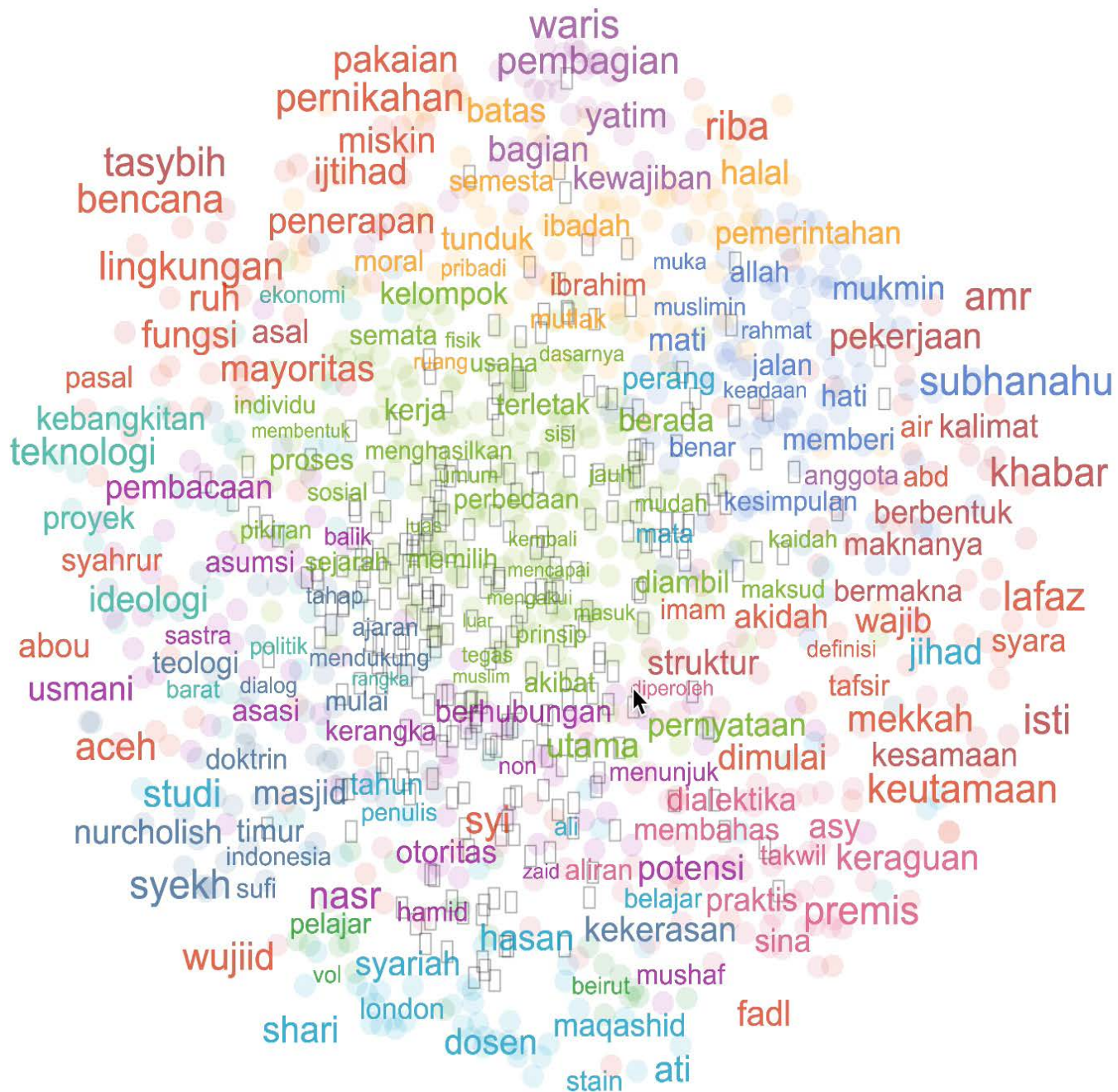
Select a Version:

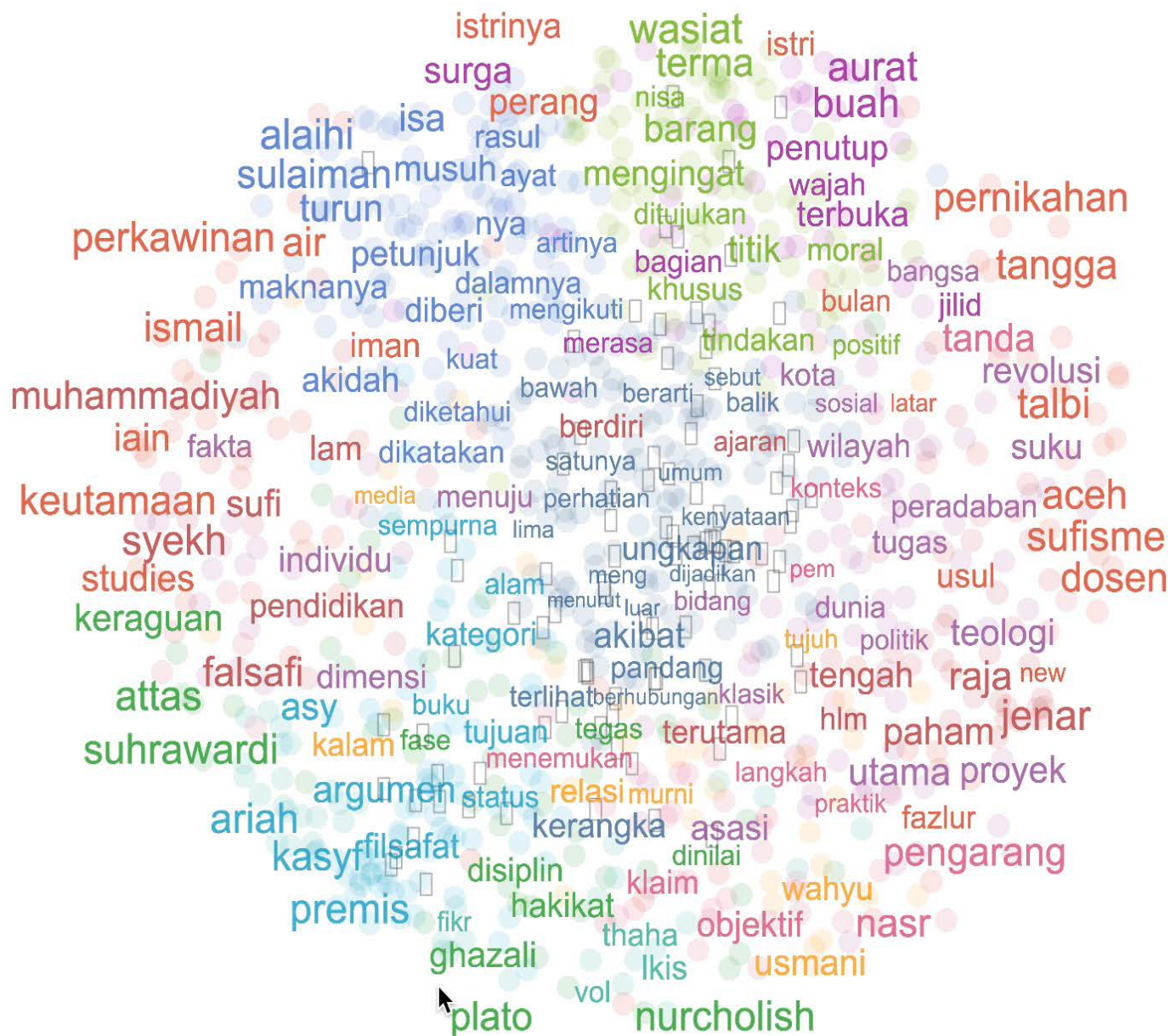
Run 5 - Sun, 11 Nov 2018 12:16:18 GMT

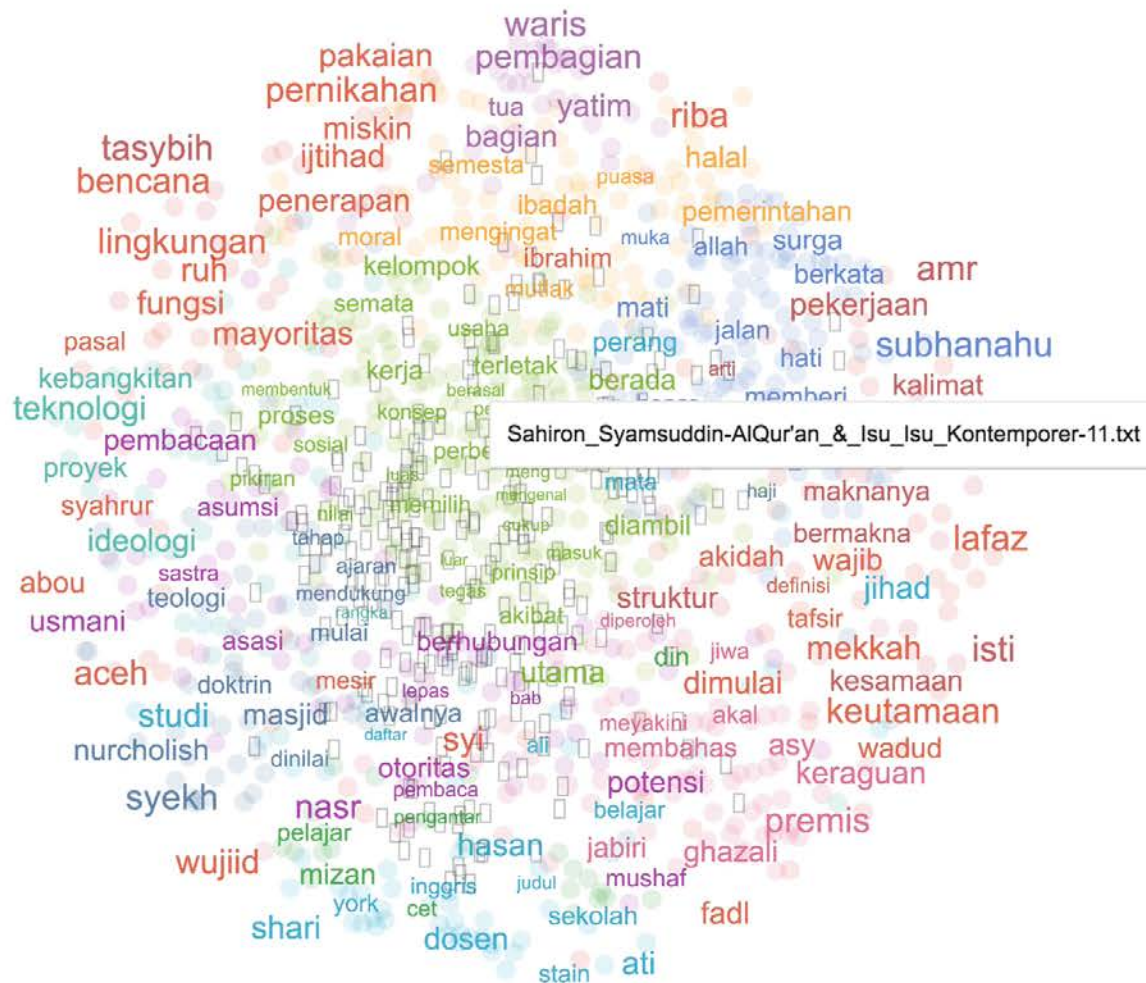
Refresh


















Elements

-  component
-  KM Entity
-  KM Path
-  Pins
-  Users

Connections

-  is
-  role
-  suggestion



Thank you!

David Vishanoff
vishanoff@ou.edu

Dave King
king@exaptive.com