Ian vs THE MACHINE

Using a PubMed Author Disambiguation Tool to Identify Publications

Mahria Lebow
Digital Projects Librarian, Kaiser Permanente Center for Health Research
PNC/MLA October 2016

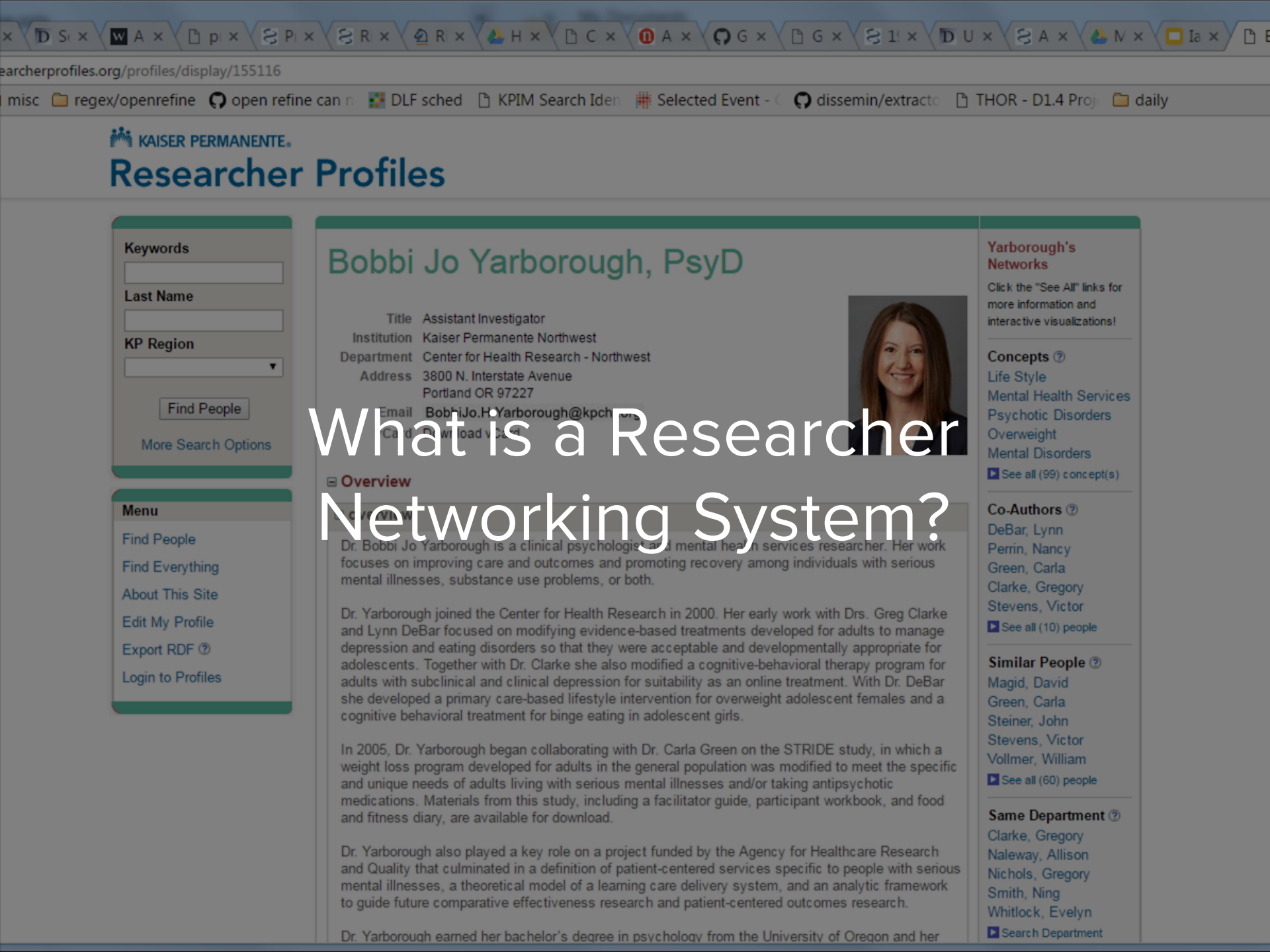
What I will touch on today:

What is a Researcher Networking System

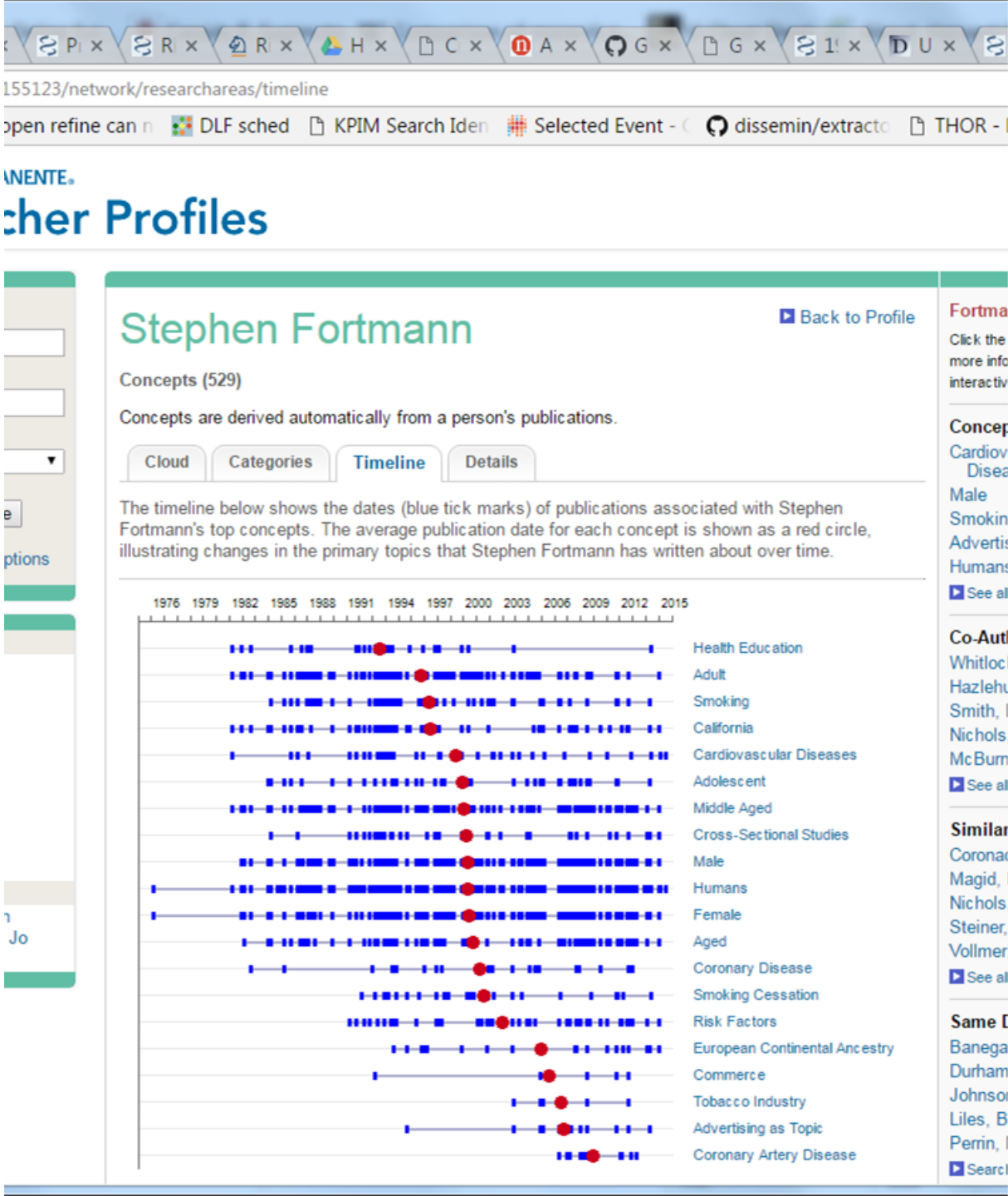
PubMed Name Disambiguation Engine

Why Ian vs The Machine

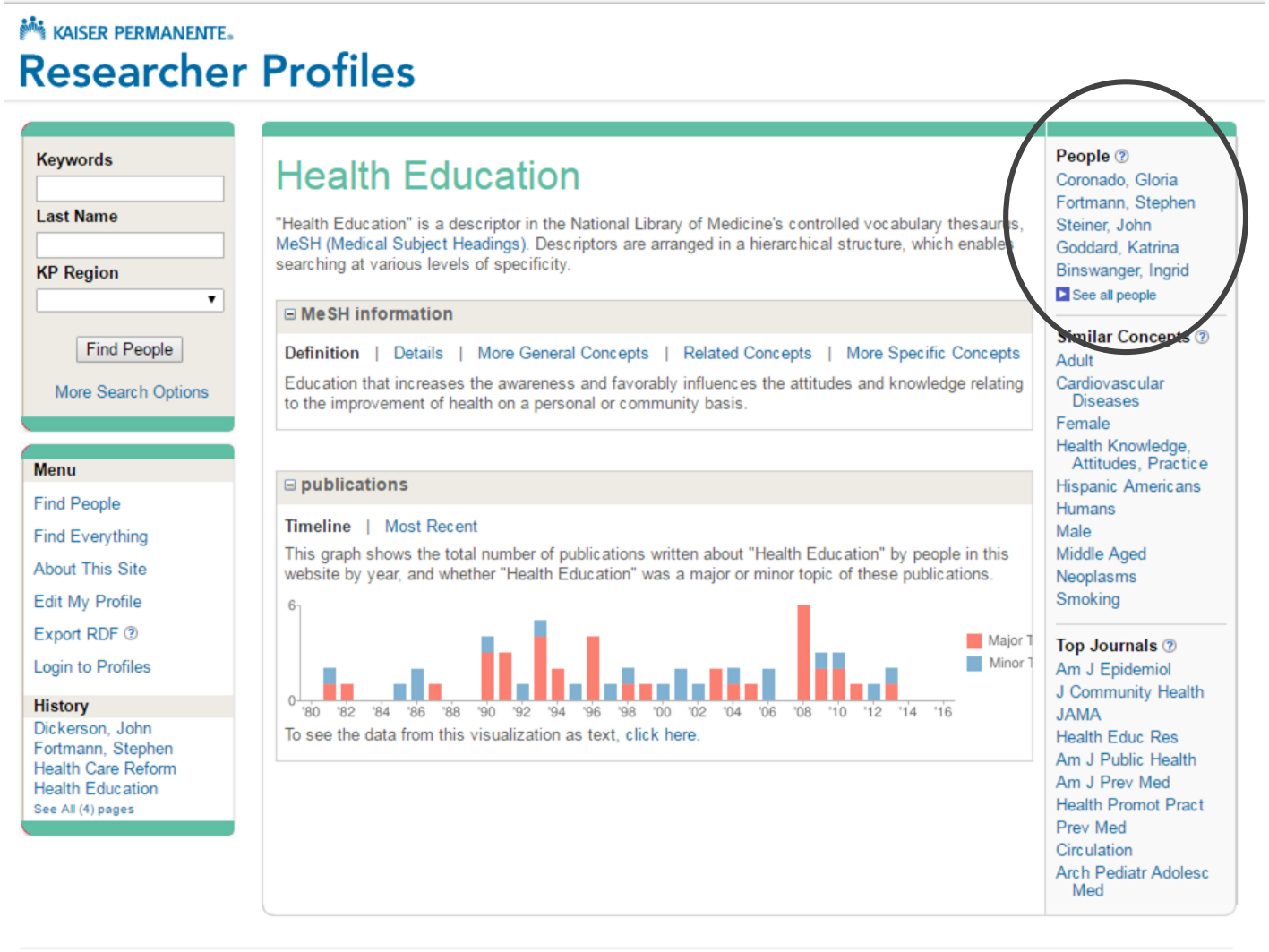
What is a Researcher Networking System?



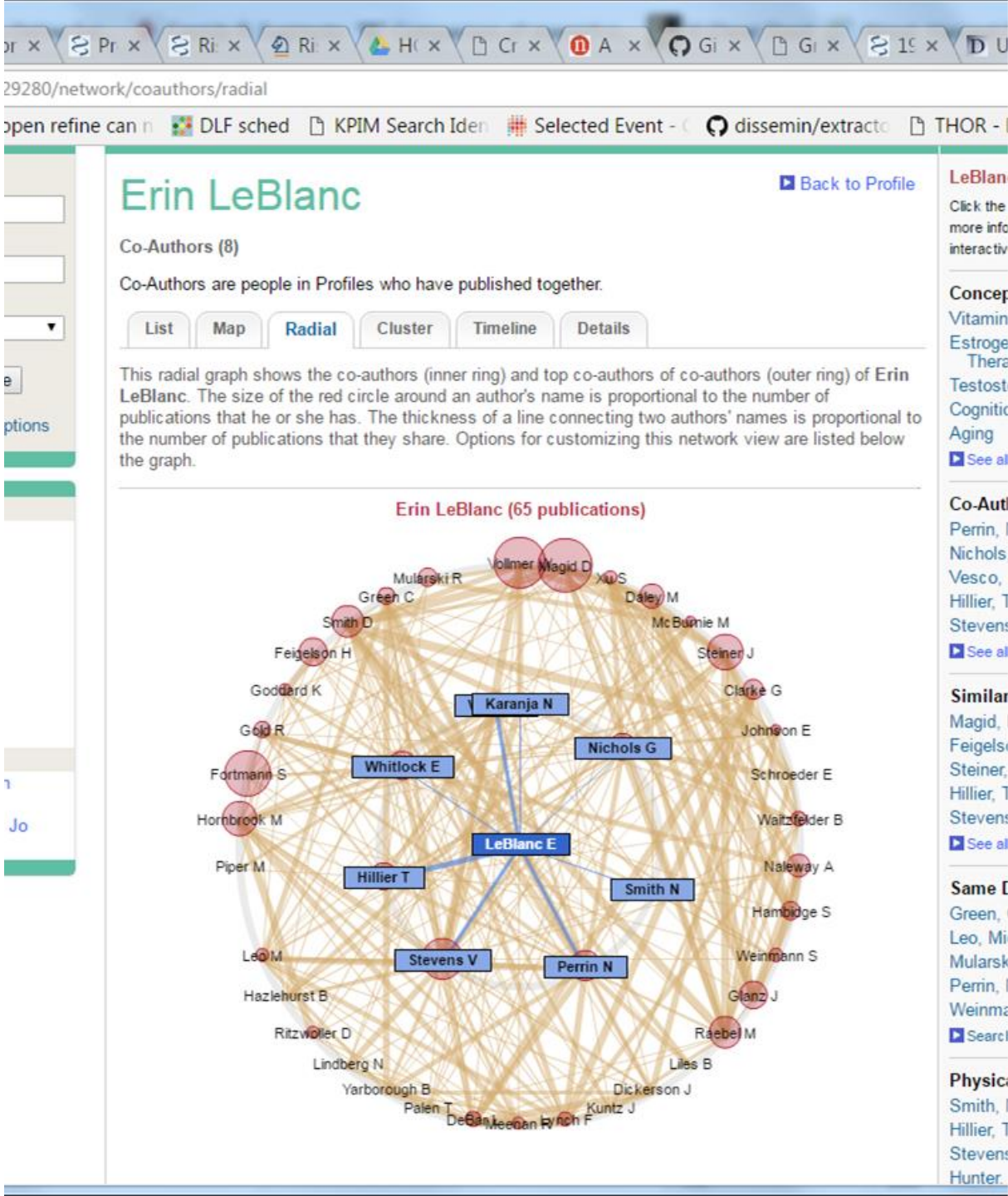
Concepts Derived From Research



Concepts Derived From Research



Co-author Networks



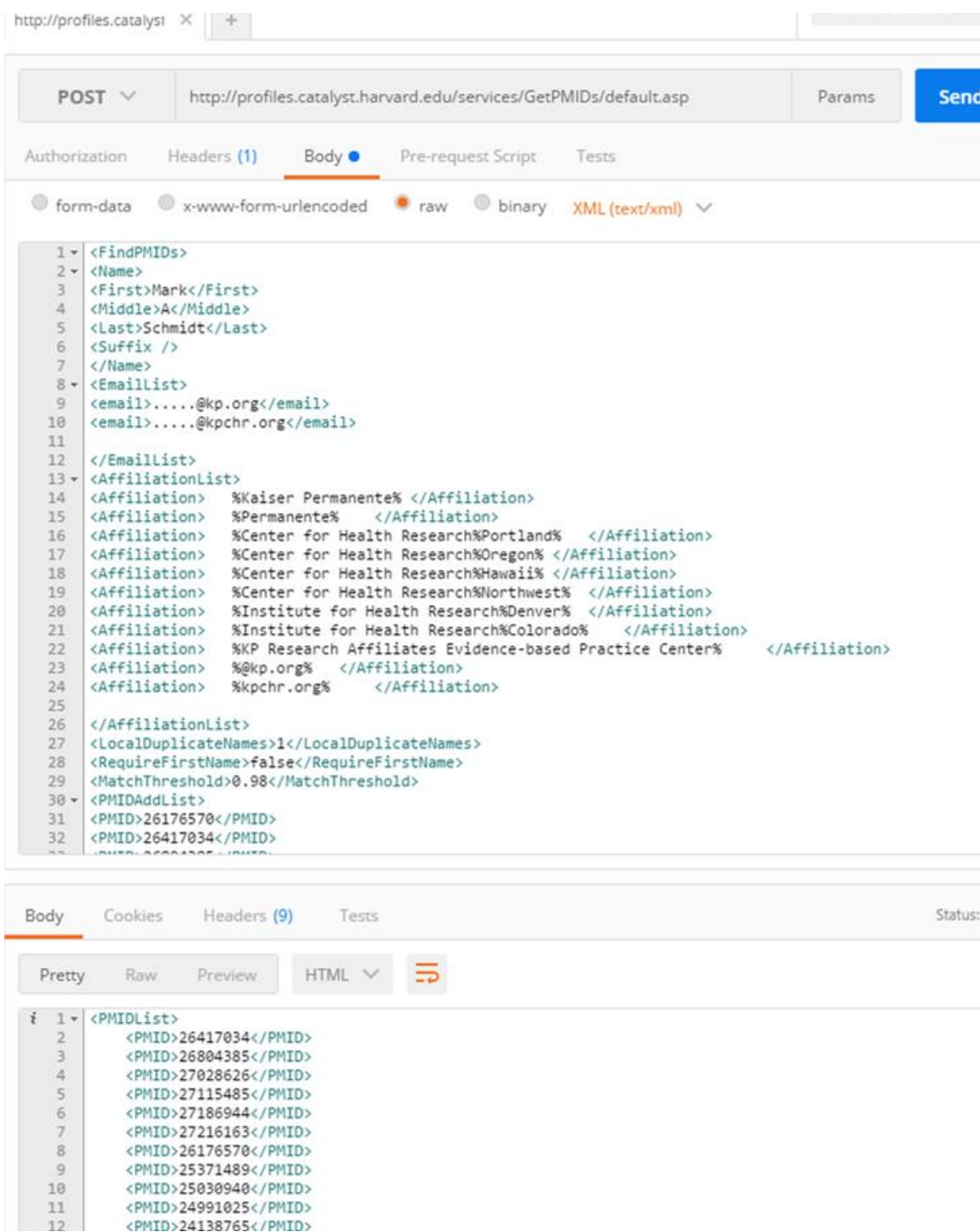
What is a name disambiguation engine?

CC by 2.0 Tiago Miranda, text was added by Mahria

Who is Ian?

Disclaimer: Ian has seen this presentation and approved the content

How does it work in Profiles RNS?



Mark M. Schmidt

- Mark Schmidt
- Kaiser Permanente
- Pregnancy (SU)
- Risk factors (SU)
- Diabetes (SU)
- Hawaii/Epidemiology (SU)
- Hillier, T (co-author)
- Mullen, Judith (co-author)

Mark A. Schmidt

- Mark Schmidt
- Kaiser Permanente
- Cohort studies (SU)
- Hep B (SU)
- Vaccines (SU)
- Vijayadeva, V (co-author)
- Trinacty, Connie (co-author)
- J Viral Hepat (journal)

Methods

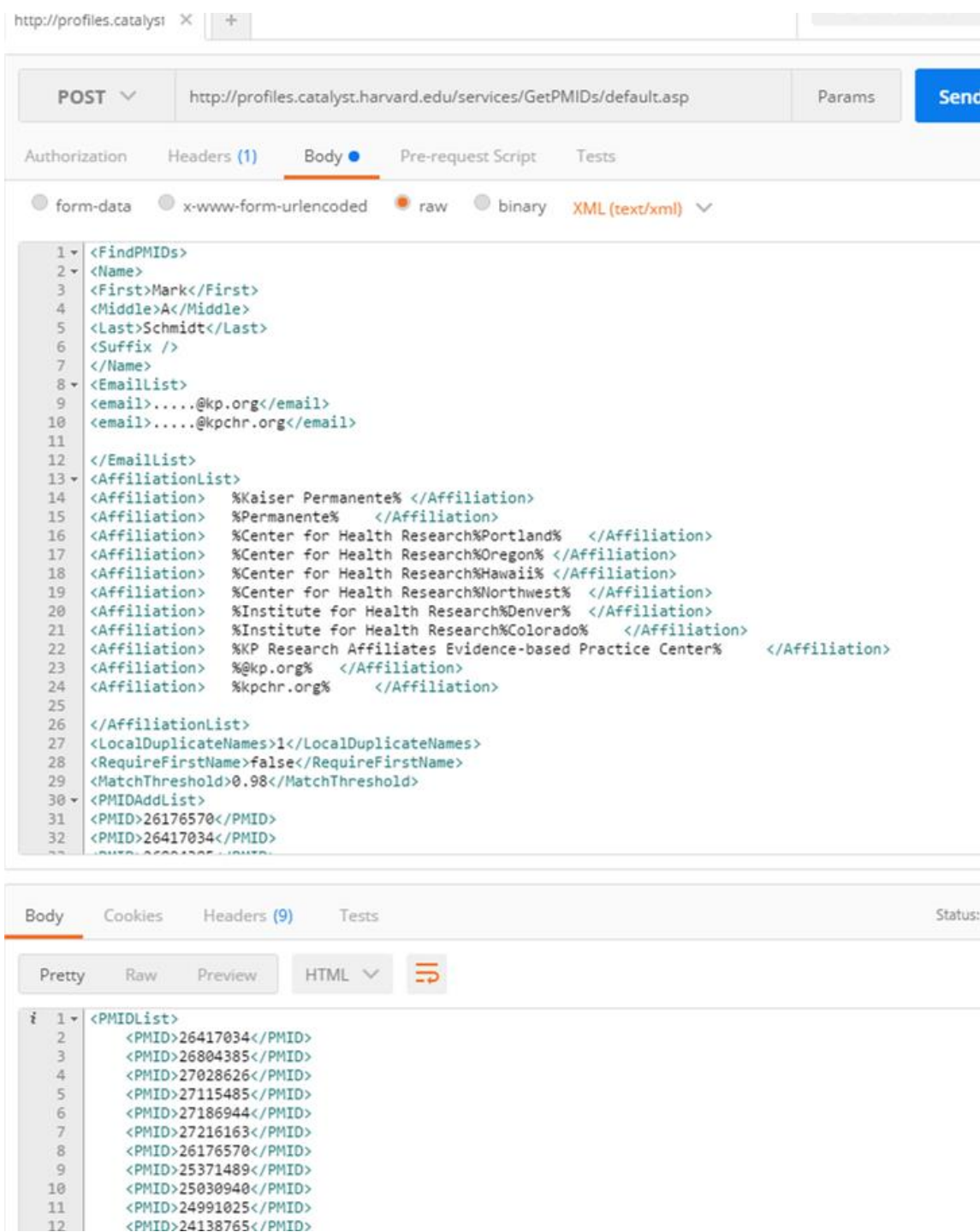
Date range: November 2015 through August 2016
Number of publications added by Ian to KPPL: 1592, 684 unique KP authors

Full name	Rank	# of pubs in PM with last name match	Mahria's difficulty rating
Steven Fortmann	Senior	306	Easiest
Erin LeBlanc	Senior	5957	Easy
Gregory Clarke	Senior	26426	Hard
Bobbi Jo Yarborough	Junior	210	Harder
Jessica Hunter	Junior	20896	Hardest
Mark Schmidt	Junior	55721	Hardest

Results

Name	# of pubs added to KPPL (Ian)	Profiles false hits (RNS)	Profiles misses (RNS)	KPPL misses (Ian)	% missed (RNS as compared to Ian)
Steven Fortmann	11	0	0 (1 epub)	0	0%
Erin LeBlanc	17	1	1 (1 epub)	0	6%
Gregory Clarke	5	0	0 (2 epubs)	0	0%
Bobbi Jo Yarborough	8	1	5	0	62%
Jessica Hunter	2	0	1 (1 epub)	N/A?	100%
Mark Schmidt	7	0	6 (1 epub)	N/A?	100%

Querying via Postman



Results - Part 2

Name	*# of pubs added to KPPL (Ian)*	# of pubs found (Postman request)	# of false hits (Postman request)	# of misses (Postman request)	# of KPPL misses (Ian)	% missed (Postman as compared to Ian)
Bobbi Jo Yarborough	*8*	8	0	0	0	*0%*
Jessica Hunter	*2*	0	0	1 (1 epub)	0	*100%*
Mark Schmidt	*7*	9	0	0 (1 epub)	2*	*0%*

Conclusion

?

References

1. Weber, G. Profiles Research Networking Software PubMed Disambiguation Service. 01/20/2012. http://profiles.catalyst.harvard.edu/Meetings/ProfilesRNSDisambiguation120120.pdf
2. TORVIK, V. I., & SMALHEISER, N. R. (2009). Author Name Disambiguation in MEDLINE. *ACM Transactions on Knowledge Discovery from Data*, *3*(3), 11. (aside: this is the AUTHOR-ity paper)
3. Profiles Research Networking Software (RNS) Disambiguation Engine Documentation. http://profiles.catalyst.harvard.edu/docs/ProfilesRNS_DisambiguationEngine.pdf. Viewed 10/4/16/.
4. Torvik, V. I., Weeber, M., Swanson, D. R., & Smalheiser, N. R. (2005). *A probabilistic similarity metric for medline records: A model for author name disambiguation*. *Journal of the American Society for Information Science and Technology*, *56*(2), 140-158. DOI: 10.1002/asi.20105