**Gerstein lab experience in text mining**

Following closely the explosion of genomics data, the number of scientific publications in the biomedical sector has seen an exponential increase. This highlighted the necessity for tools and infrastructures that would automatically process and extract the information recorded in journal articles, using a standardized vocabulary with ontological relationships. In this direction, we have developed tools for extracting and analyzing information from literature\cite{1,2,3}, including PubNet, a web-based application used to extract and integrate information from PubMed providing a graphical visualization of complex networks in order to infer functional similarities, and YeastHub, a semantic web use case for integrating data in the life sciences domain. Moreover, we proposed structures for digital publications that facilitate automated literature mining \cite{1,4,5,6,7}; analyzed trends in biomedical publications on a large scale \cite{8,9,10,11}.

Reference

1. PubNet: a flexible system for visualizing literature derived networks. Genome Biol. 2005;6(9):R80. Epub 2005 Aug 16. Douglas SM, [Montelione GT](https://www.ncbi.nlm.nih.gov/pubmed/?term=Montelione%20GT%5BAuthor%5D&cauthor=true&cauthor_uid=16168087), [Gerstein M](https://www.ncbi.nlm.nih.gov/pubmed/?term=Gerstein%20M%5BAuthor%5D&cauthor=true&cauthor_uid=16168087).

2. YeastHub: a semantic web use case for integrating data in the life sciences domain. Bioinformatics. 2005 Jun;21 Suppl 1:i85-96. [Cheung KH](https://www.ncbi.nlm.nih.gov/pubmed/?term=Cheung%20KH%5BAuthor%5D&cauthor=true&cauthor_uid=15961502)1, [Yip KY](https://www.ncbi.nlm.nih.gov/pubmed/?term=Yip%20KY%5BAuthor%5D&cauthor=true&cauthor_uid=15961502), [Smith A](https://www.ncbi.nlm.nih.gov/pubmed/?term=Smith%20A%5BAuthor%5D&cauthor=true&cauthor_uid=15961502), [Deknikker R](https://www.ncbi.nlm.nih.gov/pubmed/?term=Deknikker%20R%5BAuthor%5D&cauthor=true&cauthor_uid=15961502), [Masiar A](https://www.ncbi.nlm.nih.gov/pubmed/?term=Masiar%20A%5BAuthor%5D&cauthor=true&cauthor_uid=15961502), [Gerstein M](https://www.ncbi.nlm.nih.gov/pubmed/?term=Gerstein%20M%5BAuthor%5D&cauthor=true&cauthor_uid=15961502).

3. The tYNA platform for comparative interactomics: a web tool for managing, comparing and mining multiple networks. Bioinformatics. 2006 Dec 1;22(23):2968-70. Epub 2006 Oct 4. [Yip KY](https://www.ncbi.nlm.nih.gov/pubmed/?term=Yip%20KY%5BAuthor%5D&cauthor=true&cauthor_uid=17021160), [Yu H](https://www.ncbi.nlm.nih.gov/pubmed/?term=Yu%20H%5BAuthor%5D&cauthor=true&cauthor_uid=17021160), [Kim PM](https://www.ncbi.nlm.nih.gov/pubmed/?term=Kim%20PM%5BAuthor%5D&cauthor=true&cauthor_uid=17021160), [Schultz M](https://www.ncbi.nlm.nih.gov/pubmed/?term=Schultz%20M%5BAuthor%5D&cauthor=true&cauthor_uid=17021160), [Gerstein M](https://www.ncbi.nlm.nih.gov/pubmed/?term=Gerstein%20M%5BAuthor%5D&cauthor=true&cauthor_uid=17021160).

4. Structured digital tables on the Semantic Web: toward a structured digital literature.Mol Syst Biol. 2010 Aug 24;6:403. doi: 10.1038/msb.2010.45. [Cheung KH](https://www.ncbi.nlm.nih.gov/pubmed/?term=Cheung%20KH%5BAuthor%5D&cauthor=true&cauthor_uid=20739925), [Samwald M](https://www.ncbi.nlm.nih.gov/pubmed/?term=Samwald%20M%5BAuthor%5D&cauthor=true&cauthor_uid=20739925), [Auerbach RK](https://www.ncbi.nlm.nih.gov/pubmed/?term=Auerbach%20RK%5BAuthor%5D&cauthor=true&cauthor_uid=20739925), [Gerstein MB](https://www.ncbi.nlm.nih.gov/pubmed/?term=Gerstein%20MB%5BAuthor%5D&cauthor=true&cauthor_uid=20739925).

5. Manually structured digital abstracts: a scaffold for automatic text mining. FEBS Lett. 2008 Apr 9;582(8):1170. doi: 10.1016/j.febslet.2008.02.073. Epub 2008 Mar 6. [Seringhaus M](https://www.ncbi.nlm.nih.gov/pubmed/?term=Seringhaus%20M%5BAuthor%5D&cauthor=true&cauthor_uid=18328823), [Gerstein M](https://www.ncbi.nlm.nih.gov/pubmed/?term=Gerstein%20M%5BAuthor%5D&cauthor=true&cauthor_uid=18328823).

6. Structured digital abstract makes text mining easy. Nature. 2007 May 10;447(7141):142. Gerstein M, [Seringhaus M](https://www.ncbi.nlm.nih.gov/pubmed/?term=Seringhaus%20M%5BAuthor%5D&cauthor=true&cauthor_uid=17495904), [Fields S](https://www.ncbi.nlm.nih.gov/pubmed/?term=Fields%20S%5BAuthor%5D&cauthor=true&cauthor_uid=17495904)

7. Leveraging the structure of the Semantic Web to enhance information retrieval for proteomics. Bioinformatics. 2007 Nov 15;23(22):3073-9. Epub 2007 Oct 7. [Smith A](https://www.ncbi.nlm.nih.gov/pubmed/?term=Smith%20A%5BAuthor%5D&cauthor=true&cauthor_uid=17923450), [Cheung K](https://www.ncbi.nlm.nih.gov/pubmed/?term=Cheung%20K%5BAuthor%5D&cauthor=true&cauthor_uid=17923450), [Krauthammer M](https://www.ncbi.nlm.nih.gov/pubmed/?term=Krauthammer%20M%5BAuthor%5D&cauthor=true&cauthor_uid=17923450), [Schultz M](https://www.ncbi.nlm.nih.gov/pubmed/?term=Schultz%20M%5BAuthor%5D&cauthor=true&cauthor_uid=17923450), [Gerstein M](https://www.ncbi.nlm.nih.gov/pubmed/?term=Gerstein%20M%5BAuthor%5D&cauthor=true&cauthor_uid=17923450).

8. RNAi development. PLoS Comput Biol. 2007 Apr 27;3(4):e80. [Gerstein M](https://www.ncbi.nlm.nih.gov/pubmed/?term=Gerstein%20M%5BAuthor%5D&cauthor=true&cauthor_uid=17465677), [Douglas SM](https://www.ncbi.nlm.nih.gov/pubmed/?term=Douglas%20SM%5BAuthor%5D&cauthor=true&cauthor_uid=17465677).

9. Seeking a new biology through text mining. Cell. 2008 Jul 11;134(1):9-13. doi: 10.1016/j.cell.2008.06.029. [Rzhetsky A](https://www.ncbi.nlm.nih.gov/pubmed/?term=Rzhetsky%20A%5BAuthor%5D&cauthor=true&cauthor_uid=18614002)1, [Seringhaus M](https://www.ncbi.nlm.nih.gov/pubmed/?term=Seringhaus%20M%5BAuthor%5D&cauthor=true&cauthor_uid=18614002), [Gerstein M](https://www.ncbi.nlm.nih.gov/pubmed/?term=Gerstein%20M%5BAuthor%5D&cauthor=true&cauthor_uid=18614002).

10. The spread of scientific information: insights from the web usage statistics in PLoS article-level metrics. PLoS One. 2011;6(5):e19917. doi: 10.1371/journal.pone.0019917. Epub 2011 May 16. [Yan KK](https://www.ncbi.nlm.nih.gov/pubmed/?term=Yan%20KK%5BAuthor%5D&cauthor=true&cauthor_uid=21603617)1, [Gerstein M](https://www.ncbi.nlm.nih.gov/pubmed/?term=Gerstein%20M%5BAuthor%5D&cauthor=true&cauthor_uid=21603617).

11. Getting started in text mining: part two. PLoS Comput Biol. 2009 Jul;5(7):e1000411. doi: 10.1371/journal.pcbi.1000411. Epub 2009 Jul 31. [Rzhetsky A](https://www.ncbi.nlm.nih.gov/pubmed/?term=Rzhetsky%20A%5BAuthor%5D&cauthor=true&cauthor_uid=19649304)1, [Seringhaus M](https://www.ncbi.nlm.nih.gov/pubmed/?term=Seringhaus%20M%5BAuthor%5D&cauthor=true&cauthor_uid=19649304), [Gerstein MB](https://www.ncbi.nlm.nih.gov/pubmed/?term=Gerstein%20MB%5BAuthor%5D&cauthor=true&cauthor_uid=19649304).