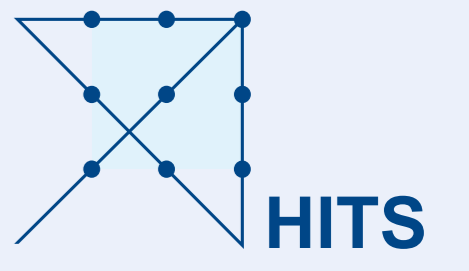


Visual Error Analysis for Entity Linking

Benjamin Heinzerling^{1,2} and Michael Strube²

¹Research Training Group AIPHES, ²Heidelberg Institute for Theoretical Studies

Heidelberg Institute for
Theoretical Studies



Entity Linking (EL)



EL system needs to:

- ▶ Detect mentions
- ▶ Link mentions to knowledge base
- ▶ Cluster mentions with no entry in knowledge base (NILs)

The Problem

▶ EL systems produce output like this:

doc id	start	end	link id	conf.	type
AFP_ENG_20090605.0363	903	914	E0455080	1.0	PER
AFP_ENG_20090605.0363	1021	1025	E0360581	1.0	PER
AFP_ENG_20091014.0713	73	80	NIL0011	1.0	LOC
AFP_ENG_20091014.0713	97	102	E0162920	1.0	PER
AFP_ENG_20091014.0713	191	198	NIL0165	1.0	GPE

▶ ... and get evaluated like this:

tp	fp	fn	precis	recall	fscore	measure
1696	1642	1608	0.508	0.513	0.511	link match
1646	774	1285	0.680	0.562	0.615	nil match
3342	2416	2893	0.580	0.536	0.557	all match
4700	1058	1531	0.816	0.754	0.784	mention match
2144	797	1997	0.729	0.518	0.606	muc
3015	2742	3966	0.524	0.364	0.429	b_cubed+

- ▶ Good for comparing systems
- ▶ Less suited for understanding what a system is doing

Goals

- ▶ Help EL researchers improve their systems
- ▶ Concisely visualize gold annotations and system output
- ▶ Make it easy to find:
 - ▷ Mention detection errors
 - ▷ Linking and NIL clustering errors
 - ▷ Errors in gold annotations

The Visual Entity Explorer (VEX)

AFP_ENG_20090605.0363 <DOC id="AFP_ENG_20091117.0647" type="story" > <HEADLINE> Microsoft co-founder Paul Allen diagnosed with cancer </HEADLINE> <DATELINE> . washington , Nov 16, 2009 (AFP) </DATELINE> <TEXT> <P> Billionaire Paul Allen , who founded US software giant Microsoft with Bill Gates in 1975, has been diagnosed with cancer; technology blogs reported on Monday. </P> <P> Allen , 56, one of the wealthiest men in the world, has non-Hodgkin's lymphoma, CNET News and other technology news sites said, quoting a letter from Allen 's sister, Jody Allen . </P> <P> "He received the diagnosis early this month and has begun chemotherapy," Jody Allen said in the letter to employees of her brother's company, Vulcan Inc . </P> <P> "Doctors say he has diffuse large B-cell lymphoma, a relatively common form of lymphoma," said Jody Allen , the chief executive of Vulcan Inc. </P> <P> She recalled that her brother survived another bout with cancer 25 years ago. </P> <P> Paul is feeling OK and remains upbeat. He continues to work and he has no plans to change his role at Vulcan , " she said. </P> <P> Allen left Microsoft in 1983 and is the founder and chairman of Vulcan Inc . and the chairman of Charter Communications . </P> <P> He owns the National Football League team the Seattle Seahawks and the Portland Trail Blazers of the National Basketball Association and is a part owner of the Major League Soccer team the Seattle Sounders . </P> <P> Allen is also a leading philanthropist with donations totaling nearly one billion dollars. </P> </TEXT> </DOC>

- Gold entities
 - Vulcan Inc.
 - Charter Communications
 - Seattle Sounders FC
 - Paul Allen
 - Portland, Oregon
 - Seattle
- TP entities
 - Vulcan Inc.
 - Charter Communications
 - Paul Allen
 - Bill Gates
 - National Football League
 - National Basketball Association
- FP entities
 - Cortland, New York
 - Vulcan Inc.
 - Allen, Texas
 - Major League Soccer
 - Paul the Apostle
 - United States

Example 1

pany, [Vulcan Inc.] . </P> <P> "Doctors say he has diffuse homa," said Jody Allen , the chief executive of [Vulcan Inc.] t with cancer 25 years ago. </P> <P> " Paul is feeling OK and s to change his role at [Vulcan Inc.]" she said. </P> <P> Allen left [Vulcan Inc.] and the chairman of Charter Communicati

Example 2

ama on the national political stage by giving the then- Illinois senator the vote speaking role at the 2004 Democratic National Convention . </P> Heinz -- the widow of Sen. John Heinz , heir to the Heinz ketchup fortune aid she found out in late September that she had cancer in her left breast after ng her annual mammogram. Heinz was bon to Portuguese parents in what ow Mozambique . </P> <P> In early October, she underwent lumpectomies on 1 breasts at a Washington hospital after doctors also discovered what they

Visualizing Entities

0091117.0647" type="story" > <HEADLINE> Microsoft co-founder Paul Allen </HEADLINE> <DATELINE> . washington , Nov 16, 2009 (AFP) </DATELINE> Paul Allen , who founded US software giant Microsoft with Bill Gates in sed with cancer, technology blogs reported on Monday. </P> <P> Allen , 56, one i the world, has non-Hodgkin's lymphoma, CNET News and other technology g a letter from Allen 's sister, Jody Allen . </P> <P> "He received the diagnosis is begun chemotherapy," Jody Allen said in the letter to employees of her ulcan Inc . </P> <P> She recalled that her brother survived another bout jo. </P> <P> " Paul is feeling OK and remains upbeat. He continues to work and nge his role at Vulcan , " she said. </P> <P> Allen left Microsoft in 1983 and an of Vulcan Inc . and the chairman of Charter Communications . </P> Allen is also a leading philanthropist with donations totaling nearly one billion

(a) Complete graph

(b) Euclidean minimum spanning tree

Using VEX



Web service at <http://cosyne.h-its.org/vex>

Source code at <http://www.github.com/noutenki/vex>

Questions or comments: benjamin.heinzerling@h-its.org



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Acknowledgements: This work has been supported by the German Research Foundation as part of the Research Training Group "Adaptive Preparation of Information from Heterogeneous Sources" (AIPHES) under grant No. GRK 1994/1, and partially funded by the Klaus Tschira Foundation, Heidelberg, Germany.