ChemSpot Evaluation Results as of 2014-01-02

SCAI [1]

Link: http://www.scai.fraunhofer.de/en/business-research-areas/bioinformatics/research-development/information-extraction-semantic-text-analysis/named-entity-recognition/chem-corpora.html

Table 1.1: Evaluation.

ChemSpot	v.2.0	v.1.6	v.1.5	OSCAR 4.1
True Positives:	1012	886	867	908
False Positives:	344	265	264	468
False Negatives:	194	320	339	298
Precision:	74.63 %	**76.98 %**	76.66 %	65.99 %
Recall:	**83.91 %**	73.47 %	71.89 %	75.29 %
F1 Score:	**79.00 %**	75.18 %	74.20 %	70.33 %

Table 1.2: Evaluation by Prediction Type (ChemSpot 2.0).
This evaluation is solely based on the predicted type. That is, the number of chemicals that were classified as belonging to a particular type and the percentage of these that are correct.

	TP	All	Precision
ABBREVIATION	102	143	71.33%
MULTIPLE	0	3	0.00%
FAMILY	99	143	69.23%
FORMULA	54	97	55.67%
IDENTIFIER	10	11	90.91%
SYSTEMATIC	473	625	75.68%
TRIVIAL	274	334	82.04%

Table 1.3: Evaluation by Goldstandard Type (ChemSpot 2.0).
This means that the goldstandard for this corpus also contains the chemical type for all annotations. A true positive is an annotation where the chemical is an exact match and the type was identified correctly. Note that this is a more strict evaluation than the one in Table 1.1 where the annotation only needs to be an exact match (and the predicted type does not matter).

	TP	FP	FN	Precision	Recall	F1 Score
ABBREVIATION	94	49	67	65.73 %	58.39 %	61.84 %
FAMILY	50	93	49	34.97 %	50.51 %	41.32 %
FORMULA	36	61	13	37.11 %	73.47 %	49.32 %
SYSTEMATIC	336	289	147	53.76 %	69.57 %	60.65 %
TRIVIAL	240	94	155	71.86 %	60.76 %	65.84 %
UNKNOWN	0	0	19	0.00 %	0.00 %	0.00 %
ALL	756	586	450	56.33 %	62.69 %	59.34 %

CRAFT [2]

Link: http://bionlp-corpora.sourceforge.net/CRAFT/

Table 2.1: Evaluation.

ChemSpot	v.2.0	v.1.6	v.1.5	OSCAR 4.1
True Positives:	2993	2399	2360	2702
False Positives:	4256	3313	3041	4767
False Negatives:	4259	5298	4892	4823
Precision:	41.29 %	42.00 %	**43.70 %**	36.18 %
Recall:	**41.27 %**	31.17 %	32.54 %	35.91 %
F1 Score:	**41.28 %**	35.78 %	37.30 %	36.04 %

Table 2.2: Evaluation by Prediction Type (ChemSpot 2.0).

	TP	All	Precision
ABBREVIATION	315	870	36.21%
FAMILY	405	540	75.00%
FORMULA	452	806	56.08%
IDENTIFIER	0	151	0.00%
MULTIPLE	0	68	0.00%
SYSTEMATIC	940	3454	27.21%
TRIVIAL	881	1360	64.78%

Table 2.3: Evaluation of normalization.

ChemSpot	v.2.0	v.1.6	v.1.5	OSCAR 4.1
Entities Total:	2993	2399	2360	2702
Entities Normalized:	2083	2037	1948	820
Normalized Correct:	1336	1213	1135	365
Percent Correct (All):	44.64 %	**50.56 %**	48.09 %	13.51 %
Precision:	**64.14 %**	59.55 %	58.26 %	44.51 %
Recall:	69.60 %	**84.91 %**	82.54 %	30.35 %
F1 Score:	66.76 %	**70.00 %**	68.31 %	36.09 %

NACTEM [3]

Link: http://www.nactem.ac.uk/metabolite-corpus/

Table 3.1: Evaluation.

ChemSpot	v.2.0	v.1.6	v.1.5	OSCAR 4.1
True Positives:	1098	1006	1017	1050
False Positives:	2852	2260	2184	2688
False Negatives:	1030	1576	1122	1146
Precision:	27.80 %	30.80 %	**31.77 %**	28.09 %
Recall:	**51.60 %**	38.96 %	47.55 %	47.81 %
F1 Score:	36.13 %	34.40 %	**38.09 %**	35.39 %

Table 3.2: Evaluation by Prediction Type (ChemSpot 2.0).

	TP	All	Precision
ABBREVIATION	117	315	37.14%
FAMILY	87	541	16.08%
FORMULA	31	271	11.44%
IDENTIFIER	0	16	0.00%
MULTIPLE	1	10	10.00%
SYSTEMATIC	408	1723	23.68%
TRIVIAL	454	1074	42.27%

DDI TRAINING [4]

Link: http://labda.inf.uc3m.es/doku.php?id=en:labda_ddicorpus

Table 4.1: Evaluation.

ChemSpot	v.2.0	v.1.6	v.1.5	OSCAR 4.1
True Positives:	10481	9917	9822	8917
False Positives:	2163	1641	1612	1917
False Negatives:	4221	4823	4880	5848
Precision:	82.89 %	85.80 %	**85.90 %**	82.31 %
Recall:	**71.29 %**	67.28 %	66.81 %	60.39 %
F1 Score:	**76.65 %**	75.42 %	75.16 %	69.67 %

Table 4.2: Evaluation by Prediction Type (ChemSpot 2.0).

	TP	All	Precision
ABBREVIATION	183	257	71.21%
FAMILY	696	934	74.52%
FORMULA	19	162	11.73%
IDENTIFIER	11	17	64.71%
MULTIPLE	1	8	12.50%
SYSTEMATIC	2680	3783	70.84%
TRIVIAL	6902	7483	92.24%

CHEBI PATENT [5]

Link: http://chebi.cvs.sourceforge.net/viewvc/chebi/chapati/patentsGoldStandard/

Table 5.1: Evaluation.

ChemSpot	v.2.0	v.1.6	v.1.5	OSCAR 4.1
True Positives:	4785	3417	3240	4106
False Positives:	12034	8190	7602	11518
False Negatives:	2935	10150	4480	3940
Precision:	28.45 %	29.44 %	**29.88 %**	26.28 %
Recall:	**61.98 %**	25.19 %	41.97 %	51.03 %
F1 Score:	**39.00 %**	27.15 %	34.91 %	34.69 %

Table 5.2: Evaluation by Prediction Type (ChemSpot 2.0).

	TP	All	Precision
ABBREVIATION	180	710	25.35%
FAMILY	678	2420	28.02%
FORMULA	251	1840	13.64%
IDENTIFIER	3	601	0.50%
MULTIPLE	1	74	1.35%
SYSTEMATIC	2442	7978	30.61%
TRIVIAL	1232	3194	38.57%
UNKNOWN	0	2	0.00%

Table 5.3: Evaluation of Normalization.

ChemSpot	v.2.0	v.1.6	v.1.5	OSCAR 4.1
Entities Total:	4785	3417	3240	4106
Entities Normalized:	2002	2117	1806	873
Normalized Correct:	1343	1359	1049	519
Percent Correct (All):	28.07 %	39.77 %	32.38 %	12.64 %
Precision:	**67.08 %**	64.19 %	58.08 %	59.45 %
Recall:	41.84 %	**61.95 %**	55.74 %	21.26 %
F1 Score:	51.54 %	**63.05 %**	56.89 %	31.32 %

CHEMDNER TRAINING [6]

Link: http://www.biocreative.org/news/corpora/bc-iv-chemdner-corpus/

Table 6.1: Evaluation.

ChemSpot	v.2.0	v.1.6	v.1.5*	OSCAR 4.1
True Positives:	17706	13407	n/a	20726
False Positives:	15970	13342	n/a	10007
False Negatives:	11449	15941	n/a	8752
Precision:	52.58 %	50.12 %	n/a	**67.44 %**
Recall:	60.73 %	45.69 %	n/a	**70.31 %**
F1 Score:	56.36 %	47.80 %	n/a	**68.84 %**

**ChemSpot 1.5 does not have the CHEMDNER corpus reader and hence could not be evaluated*

Table 6.2: Evaluation by Prediction Type (ChemSpot 2.0).

	TP	All	Precision
ABBREVIATION	2790	6140	45.44%
FAMILY	2294	3449	66.51%
FORMULA	2110	3679	57.35%
IDENTIFIER	210	439	47.84%
MULTIPLE	50	92	54.35%
SYSTEMATIC	5514	12948	42.59%
TRIVIAL	5234	8095	64.66%
UNKNOWN	0	1	0.00%

Table 6.3: Evaluation by Goldstandard Type (ChemSpot 2.0).

	TP	FP	FN	Precision	Recall	F1 Score
ABBREVIATION	2307	1312	1950	61.26 %	51.55 %	55.99 %
FAMILY	2075	1312	1950	61.26 %	51.55 %	55.99 %
FORMULA	2007	1630	2296	55.18 %	46.64 %	50.55 %
IDENTIFIER	201	237	471	45.89 %	29.91 %	36.22 %
MULTIPLE	49	43	153	53.26 %	24.26 %	33.33 %
SYSTEMATIC	3934	8506	2600	31.62 %	60.21 %	41.47 %
TRIVIAL	4770	3156	3881	60.18 %	55.14 %	57.55 %
UNKNOWN	0	1	40	0.00 %	0.00 %	0.00 %
ALL	15343	18551	13552	45.27 %	53.10 %	48.87 %

References

[1] Kolářik, C., Klinger, R., Friedrich, C.M., Hofmann-Apitius, M., Fluck, J. *Chemical Names: Terminological Resources and Corpora Annotation. In Workshop on Building and evaluating resources for biomedical text mining,* 6th edition of the Language Resources and Evaluation Conference

[2] Bada, M., Eckert, M., Evans, D., Garcia, K., Shipley, K., Sitnikov, D., Baumgartner Jr., W. A., Cohen, K. B., Verspoor, K., Blake, J. A., Hunter, L. E. *Concept Annotation in the CRAFT Corpus. BMC Bioinformatics. 2012*

[3] Herrgård, M. J., Swainston, N., Dobson, P., Dunn, W. B., Arga, K. Y., Arvas, M., Büthgen, N., Borger, S., Costenoble, R., Heinemann, M., Hucka, M., Novère, N. L., Li, P., Liebermeister, W., Mo, M. L., Oliveira, A. P., Petranovic, D., Pettifer, S., Simeonidis, E., Smallbone, K., Spasíc, I.,Weichart, D., Brent, R., Broomhead, D. S., Westerhoff, H. V., Kürdar, B., Penttilä, M., Klipp, E., Palsson, B. Ø., Sauer, U., Oliver, S. G., Mendes, P., Nielsen, J. & Kell, D. B. (2008). *A consensus yeast metabolic reconstruction obtained from a community approach to systems biology*. Nature Biotechnology 26*,* 1155–1160.

[4] María Herrero-Zazo, Isabel Segura-Bedmar, Paloma Martínez, Thierry Declerck, *The DDI corpus: An annotated corpus with pharmacological substances and drug–drug interactions*, Journal of Biomedical Informatics, Volume 46, Issue 5, October 2013, 914-920

[5] A collaboration between the European Patent Office and the ChEBI team. No further reference or publication appears to exist.

[6] Krallinger, M., Leitner, F., Rabal, O., Vazquez, M., Oyarzabal, J., Valencia, A., *Overview of the chemical compound and drug name recognition (CHEMDNER) task*, Proceedings of the Fourth BioCreative Challenge Evaluation Workshop vol. 2 ,2-33