



i2b2/UTHealth NLP 2014: Wrap up

Amber Stubbs, Christopher Kotfila, Hua Xu, Ozlem Uzuner

November 14, 2014



Task 1:

De-identification



De-id annotation

- Time-consuming
- Doesn't matter if annotation is done in parallel (two annotators work separately) or in serial (first annotator works on file, then passes it to second)
 - Stubbs, A. and Uzuner, O. De-identification of Medical Records Through Annotation. Handbook of Linguistic Annotation, Nancy Ide and James Pustejovsky, eds. Forthcoming.

De-id system approaches

Rank	Team Name	Approach
1	University of Nottingham	CRF + rules
2	Harbin Institute of Technology: Shenzhen Graduate School	CRF + rules
3	University of Manchester	CRF + rules
4	Harbin Institute of Technology	CRF
5	Kaiser Permanente	Combination of existing tools + rules
6	LIMSI–CNRS	CRF + rules
7	Memorial University of Newfoundland	Bayesian HMM
8	East China Normal University	unknown
9	California State University San Marcos	Decision trees + rules
10	UC San Diego	unknown

Comparison to 2006 HIPAA categories only

Group	F-measure
Wellner	0.981
Szarvas	0.98
Aramaki	0.97
Hara	0.949
Wrenn	0.946
Guo	0.909
Guillen	0.801

2006 token-based
evaluations, best
f-measure from each
group

Group	F-measure
Nottingham	0.9757
UNIMAN	0.9668
Harbin-Grad	0.9662
Harbin	0.9536
Kaiser	0.9221
Newfoundland	0.8688
NIMSI-CNRS	0.8675
ECNU	0.8237
San Marcos	0.7255
UCSD	0.6043

2014 token-based evaluations,
best f-measure from each
group

PHI vs. non-PHI: entity-based, HIPAA categories only

Group	F-measure
Szarvas	0.967
Wellner	0.963
Aramaki	0.941
Hara	0.903
Wrenn	0.902
Guo	0.813

2006 entity-based
PHI/non-PHI
evaluations, best
f-measure from each
group

Group	F-measure
Nottingham	.9573
Harbin-Grad	.9409
UNIMAN	.9323
Harbin	.9180
Kaiser	.8429
LIMSI-CNRS	.838
Newfoundland	.7985
ECNU	.7391
San Marcos	.5733
UCSD	.4524

2014 entity-based PHI/non-PHI
evaluations, best f-measure
from each group

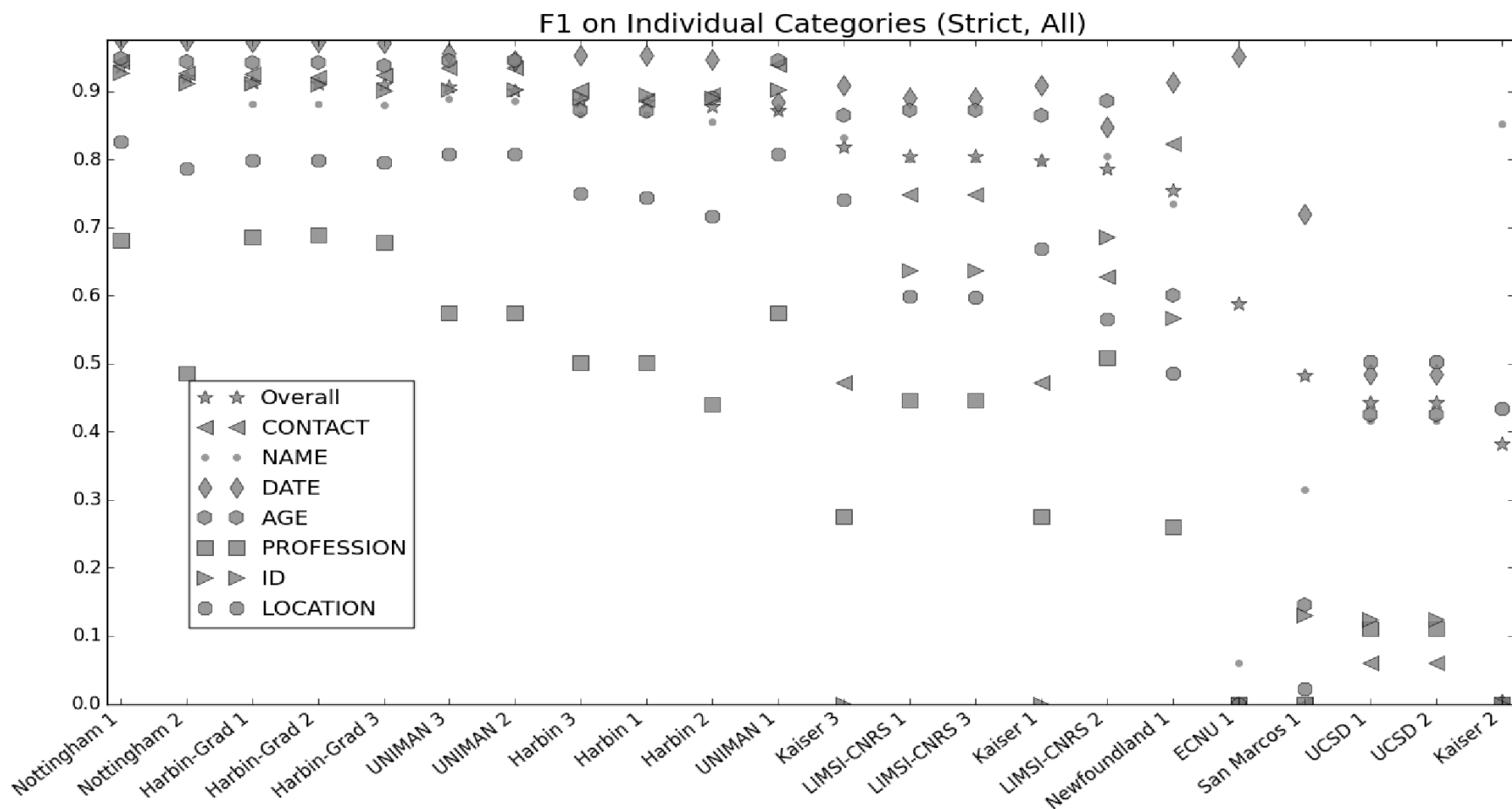


Comparison to 2006

- Overall numbers for 2014 appear slightly lower
 - 2006 data was cleaned and tokenized
 - 2014 included more types of medical records

De-id: PHI categories

3. F1 on Individual Categories (Strict, All)





De-id: categories

- Profession
 - Few examples in corpus (413 total)
- Location
 - Was the lowest-scoring category in 2006 challenge as well
 - Some hardly represented in the corpus
 - Other: 17
 - Country: 183
 - Organization: 206
 - The content of the organization tags varied widely



Task 2:

Risk factors



RF annotation

- “Light” annotation, voted gold standard: did it work?
 - Faster than exhaustive annotations, enabled more files to be annotated
 - Roughly 10min/record
 - High IAA
 - Some errors in gold standard
 - Systems achieved high scores
 - But, imperfect annotations; lack of negative examples; very skewed examples of indicators

RF: system approaches

Rank	Team name	System
1	NLM	Rules + SVM
2	Harbin Institute of Technology Shenzhen Graduate School	Rules + CRF, SSVM
3	Kaiser Permanente	Rules + ML (various)
4	Linguamatics and Northwestern University	Rules + Decision tree
5	University of Nottingham	ML (unspecified) + rules
6	The Ohio State University	Rule-based
7	TMUNSW	Rules + CRF, Naive Bayes
8	National Central University	unknown
9	University of Manchester	Rule-based
10	University of Utah	ML (various)

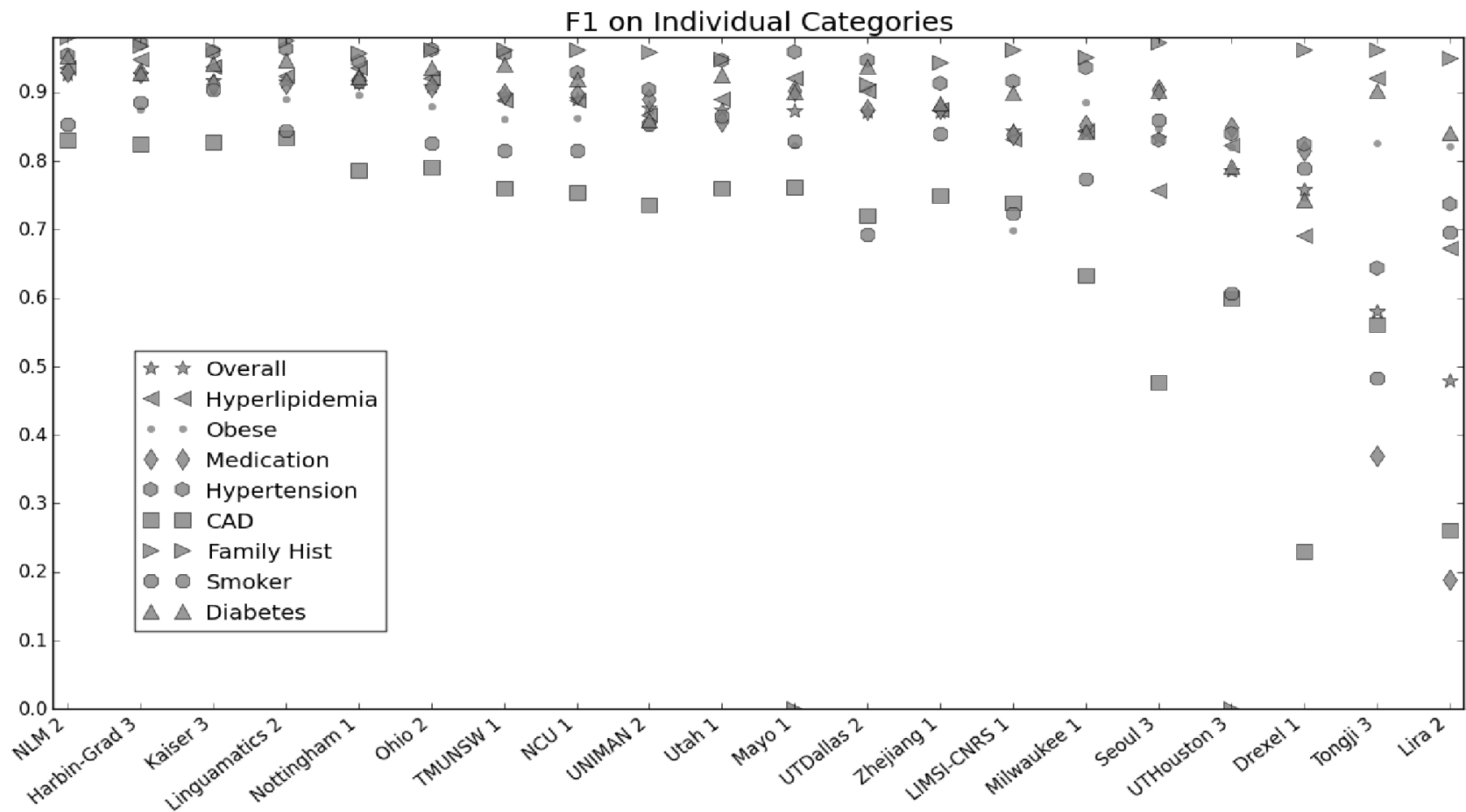


Common approaches

- Use high-recall systems, then remove inaccurate candidates
- Hybrid: ML + rules
- Use of existing systems

RF: accuracy of different risk factors

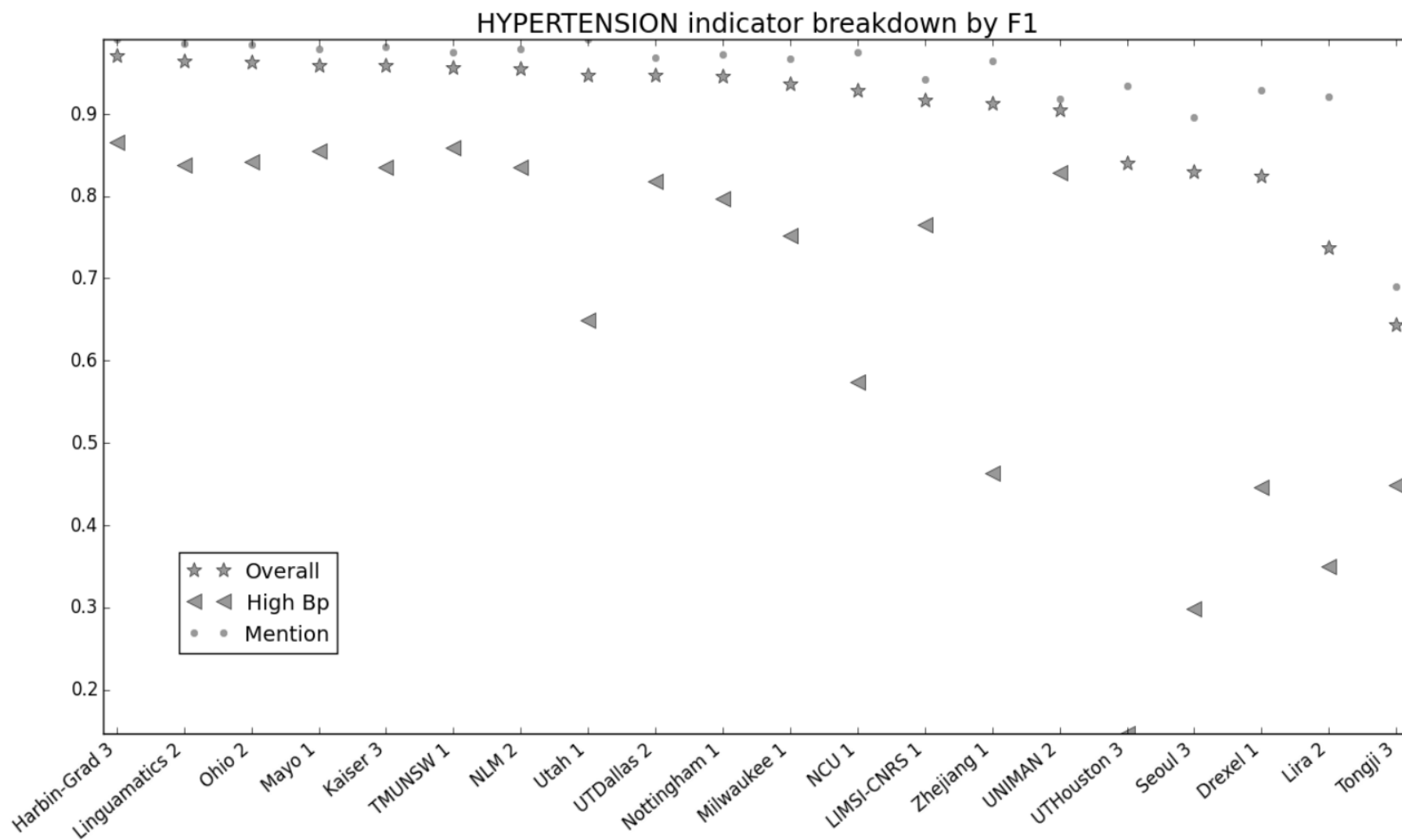
3. F1 on Individual Tag Types



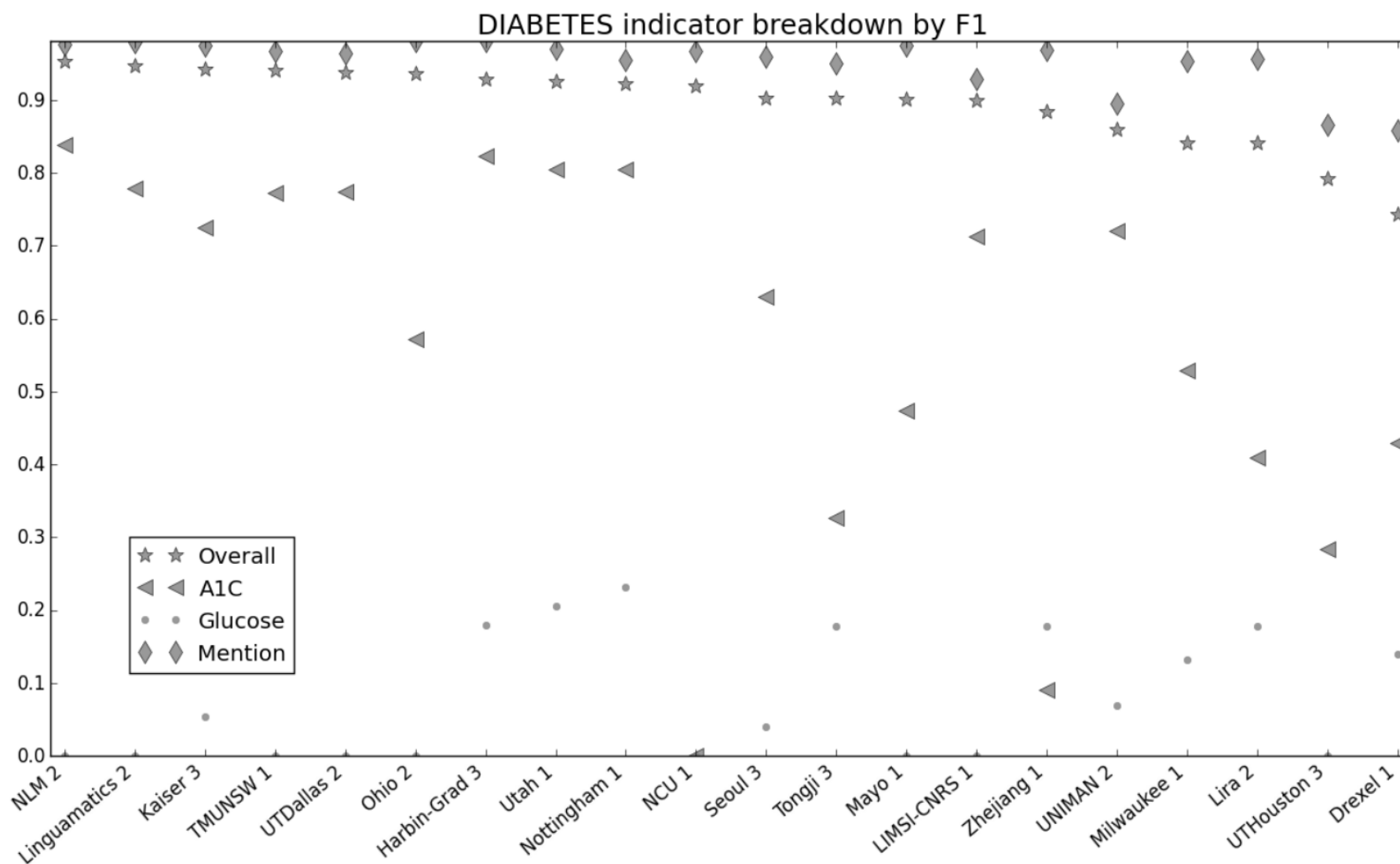
F1 measures on risk factors

System	Overall	Hyperlip.	Obese	Med	HTN	CAD	Fam Hist	Smoker	Diabetes
NLM	0.9276	0.9366	0.9298	0.9307	0.9555	0.8303	0.9805	0.8538	0.9533
Harbin-Grad	0.9268	0.9491	0.8757	0.9293	0.9715	0.8253	0.9681	0.8861	0.9291
Kaiser	0.9185	0.9375	0.9011	0.9126	0.9591	0.8284	0.963	0.9045	0.942
Lingua matics	0.9171	0.9236	0.8902	0.9138	0.9647	0.8331	0.9767	0.8441	0.9473
Nottingham	0.9156	0.9368	0.8961	0.9194	0.9455	0.787	0.9572	0.9162	0.9228
Ohio	0.9081	0.9204	0.8801	0.9092	0.963	0.7904	0.963	0.8264	0.9369
TMUNSW	0.8973	0.8896	0.862	0.8982	0.9567	0.7598	0.963	0.8148	0.9406
NCU	0.8909	0.8896	0.8624	0.8982	0.9289	0.7536	0.963	0.8148	0.9191
UNIMAN	0.8776	0.8669	0.8486	0.89	0.9047	0.7363	0.9591	0.8538	0.8603
Utah	0.8747	0.8903	0.8581	0.8585	0.9477	0.7603	0.9494	0.8655	0.9256

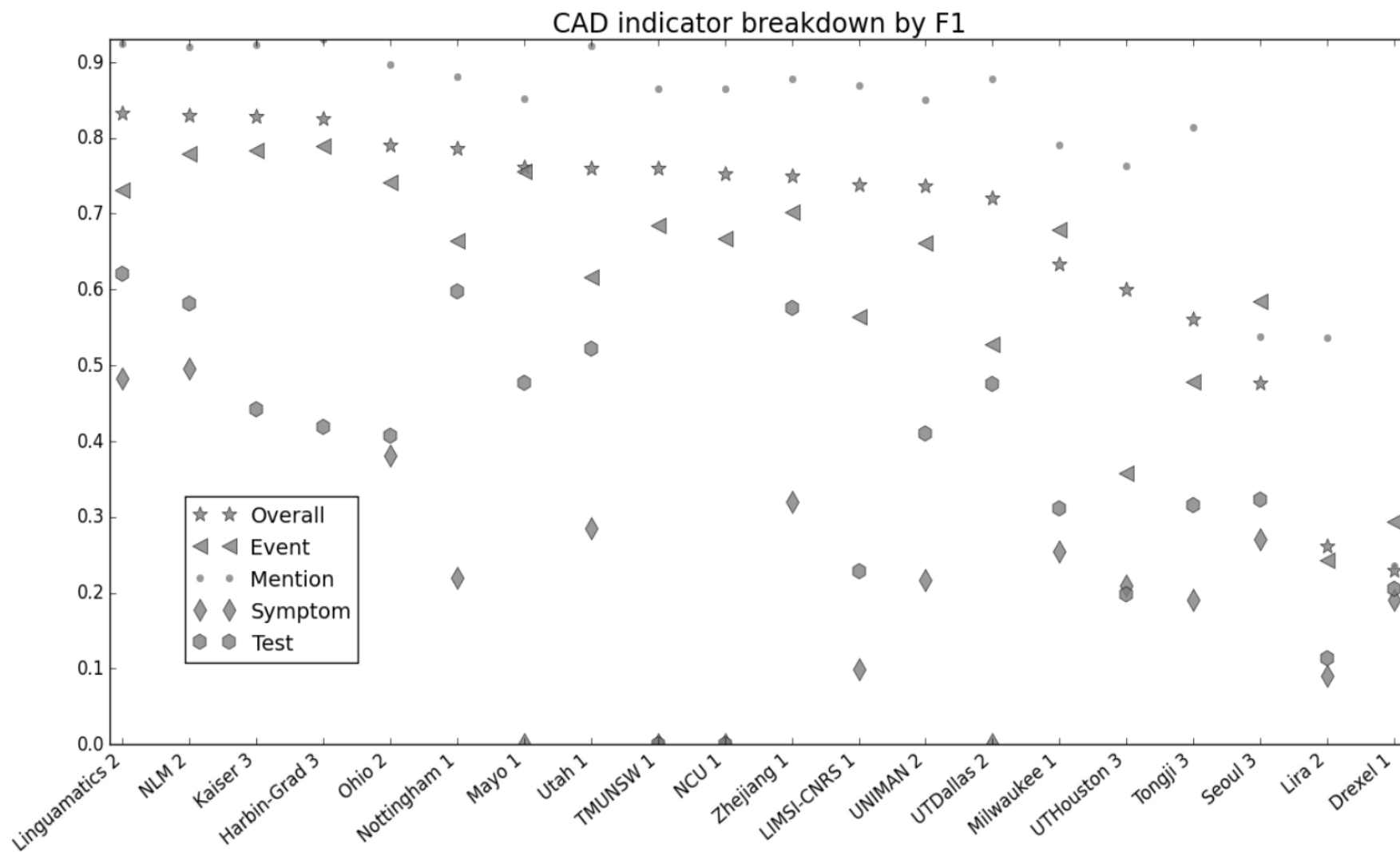
Hypertension: Indicators



Diabetes: Indicators



CAD: Indicators





Journal of Biomedical Informatics



i2b2 special issue

Deadline: January 15, 2015

All i2b2 participants are
welcome to submit