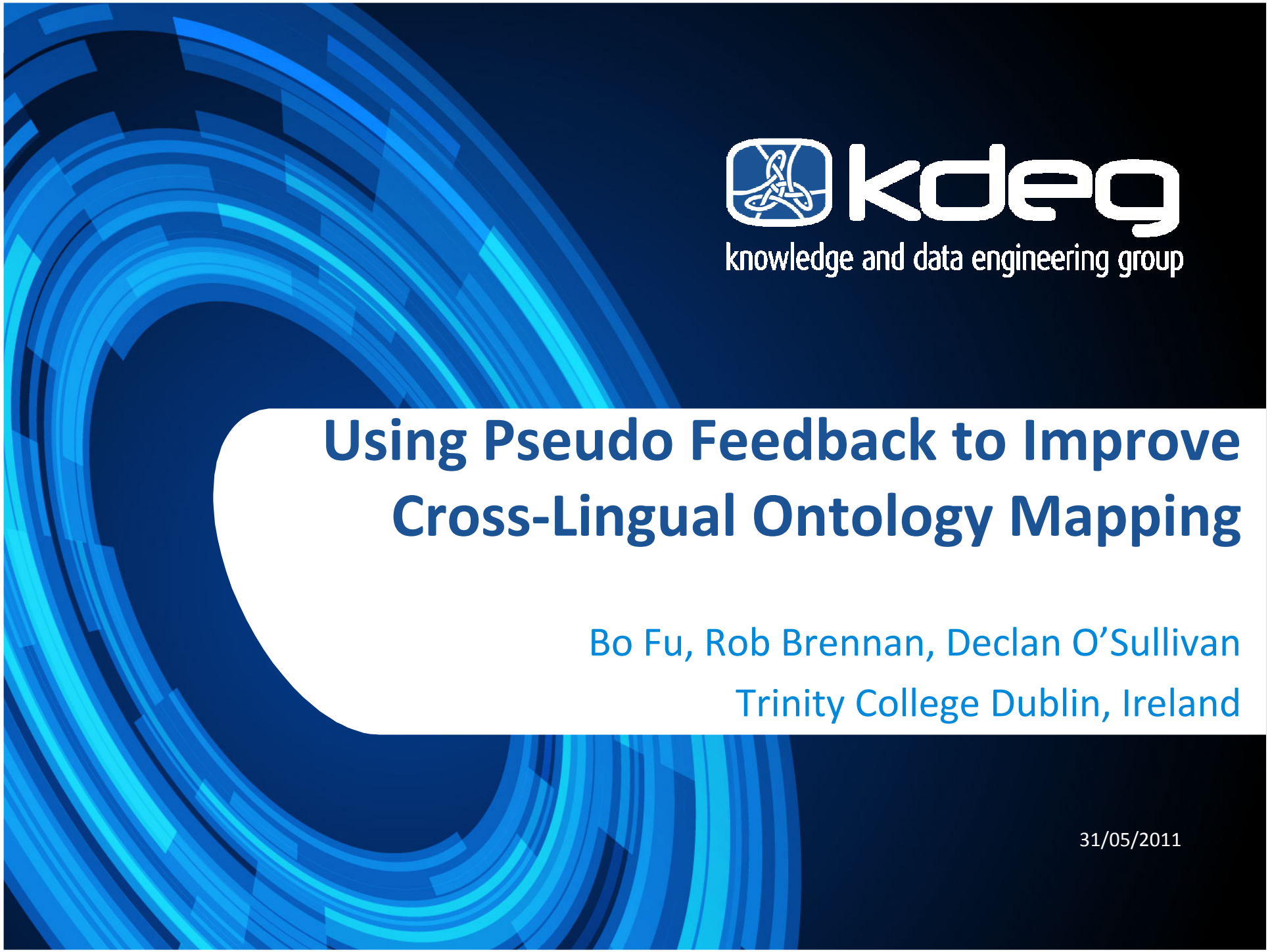


Using Pseudo Feedback to Improve Cross-Lingual Ontology Mapping

Bo Fu, Rob Brennan, Declan O'Sullivan

Trinity College Dublin, Ireland

31/05/2011

Overview

- Introduction
- Related Work
- Challenge
- Pseudo Feedback
- Experiment & Results
- Conclusions & Future Work

Introduction

- What?
  - A definition: Cross-Lingual Ontology Mapping (CLOM) refers to the process of establishing relationships among ontological resources from two or more independent ontologies where each ontology is labelled in a different natural language.
- Why?
  - Natural language independent vs. language dependent
  - Multilingual semantic web & multilingual ontologies
  - Industry requirements
  - Academia efforts & research trends
- Benefits?
  - Better digital content management
  - New ways to support semantic search
  - Enhanced personalisation and adaptation

Related Work

- CLOM Approaches:
  - Manual
  - Corpus-based
  - Via linguistic enrichment
  - Via indirect alignment
  - Translation-based

Manual CLOM

Accurate
Reliable

Time-consuming
Labour-intensive

- E.g. mapping of AGROVOC & CAT [1]

[1] Liang, A., Sini, M.: Mapping AGROVOC & the Chinese Agricultural Thesaurus: Definitions, Tools Procedures. New Review of Hypermedia & Multimedia 12(1), 51–62 (2006)

Corpus-Based CLOM

Unparallel bilingual corpora

Relies on synsets
Construction overhead
Domain coverage

- E.g. alignment of WordNet and HowNet [2]

[2] Ngai, G., Carpuat, M., Fung, P.: Identifying Concepts Across Languages: A First Step towards A Corpus-based Approach to Automatic Ontology Alignment. In: Proceedings of the 19th International Conference on Computational Linguistics, vol. 1, pp. 1–7 (2002)

CLOM via Linguistic Enrichment

Linguistic evidence
Enrichment tools [3]

Unstandardised
Difficult to build upon

- E.g. linguistically motivated mapping [4]

[3] Pazienza M. , Stellato A.: Exploiting Linguistic Resources for Building Linguistically Motivated Ontologies in the Semantic Web. In: Proceedings of OntoLex Workshop 2006: Interfacing Ontologies and Lexical Resources for Semantic Web Technologies (2006)

[4] Pazienta, M., Stellato, A.: Linguistically Motivated Ontology Mapping for the Semantic Web. In: Proceedings of the 2nd Italian Semantic Web Workshop, pp. 14–16 (2005)

CLOM via Indirect Alignment

Easily achievable

Dependent upon CLOM results
CLOM remains a challenge in itself

- E.g. mapping reuse [5]

[5] Jung, J.J., Håkansson, A., Hartung, R.: Indirect Alignment between Multilingual Ontologies: A Case Study of Korean and Swedish Ontologies. In: Håkansson, A., Nguyen, N.T., Hartung, R.L., Howlett, R.J., Jain, L.C. (eds.) KES-AMSTA 2009. LNCS (LNAI), vol. 5559, pp. 233–241. Springer, Heidelberg (2009)

Translation-Based CLOM

Available translation & MLOM tools
Feasible & scalable

?

- E.g. use of bilingual dictionaries [6], multilingual thesauri [7], off-the-shelf machine translation (MT) tools [8, 9], followed by monolingual ontology matching (MLOM) processes

[6] Zhang, X., Zhong, Q., Li, J., Tang, J., Xie, G., Li, H.: RiMOM Results for OAEI 2008. In: Proceedings of the 3rd International Workshop on Ontology Matching, pp. 182–189 (2008)

[7] Bouma, G.: Cross-lingual Ontology Alignment using EuroWordNet and Wikipedia. In: Proceedings of the 7th Conference on International Language Resources and Evaluation (LREC 2010), pp. 1023–1028 (2010) ISBN 2-9517408-6-7

[8] Wang, S., Isaac, A., Schopman, B., Schlobach, S., Van der Meij, L.: Matching Multilingual Subject Vocabularies. In: Agosti, M., Borbinha, J., Kapidakis, S., Papatheodorou, C., Tsakonas, G. (eds.) ECDL 2009. LNCS, vol. 5714, pp. 125–137. Springer (2009)

[9] Trojahn, C., Quaresma, P., Vieira, R.: An API for Multi-lingual Ontology Matching. In: Proceedings of the 7th Conference on International Language Resources and Evaluation, pp. 3830–3835 (2010) ISBN 2-9517408-6-7

The Translation Challenge in CLOM

- Motivation
  - Translation for the purpose of localisation vs. translation for the purpose of mapping
- Previous Work
  - MT tools can introduce noise into the mapping process for MLOM tools [10], where noise refers to translations that led to incorrect mapping or neglected correct mapping results
- Appropriate Translated Ontology Labels (AOLT)
  - an appropriate ontology label translation (AOLT) in the context of CLOM is one that is most likely to maximize the success of the subsequent MLOM step.
  - the translation of ontology labels refers to the translation of strings that are used to identify ontological resources in a formally defined ontology
    - e.g. the value of `rdf:ID` in `<Class rdf:ID="Thing"/>`
    - or the fragment identifier, i.e. the string after the hash sign in `<owl:Class rdf:about="http://swrc.ontoware.org/ontology#Person"/>`
    - It does not refer to the content of `rdfs:label` elements such as `<rdfs:label>Thing</rdfs:label>`.

[10] Fu, B., Brennan, R., O'Sullivan, D.: Cross-Lingual Ontology Mapping – An Investigation of the Impact of Machine Translation. In: Gómez-Pérez, A., Yu, Y., Ding, Y. (eds.) ASWC 2009. LNCS, vol. 5926, pp. 1–15. Springer, Heidelberg (2009)

Inspiration

- Relevance Feedback in Information Retrieval (IR)
  - Explicit
  - Implicit
  - Pseudo
- Use of Feedback in ontology mapping
  - Explicit feedback in MLOM [11]

[11] Duan, S., Fokoue, A., Srinivas, K.: One size does not fit all: Customizing Ontology Alignment using User Feedback. In: Patel-Schneider, P.F., Pan, Y., Hitzler, P., Mika, P., Zhang, L., Pan, J.Z., Horrocks, I., Glimm, B. (eds.) ISWC 2010, Part II. LNCS, vol. 6497, pp. 177-192. Springer, Heidelberg (2010)

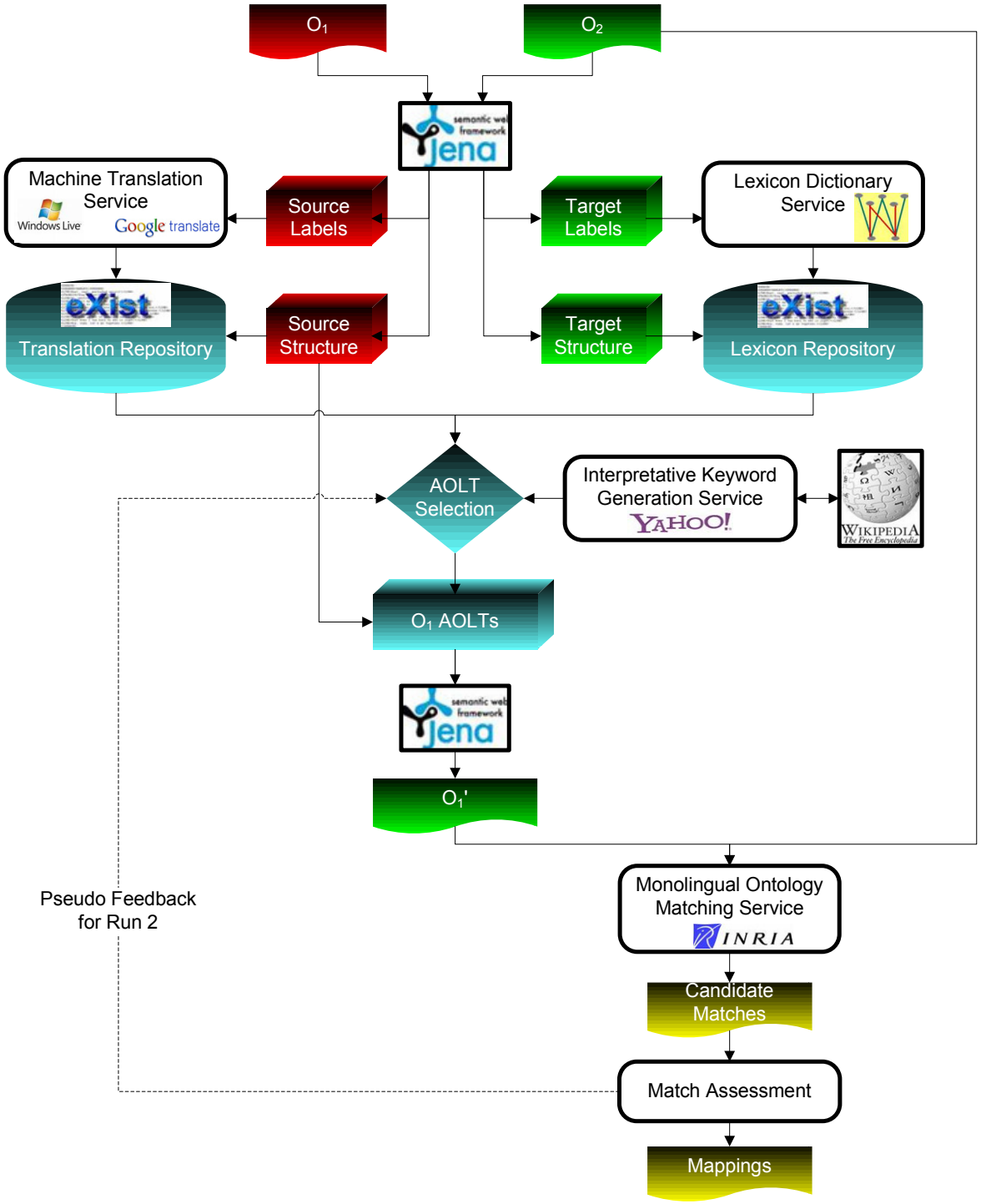
Pseudo Feedback in CLOM

Pseudo Feedback in IR

- Generate initial set of documents
- Make assumptions on their relevance
- Compute a better set of documents based on this assumption

Pseudo Feedback in CLOM

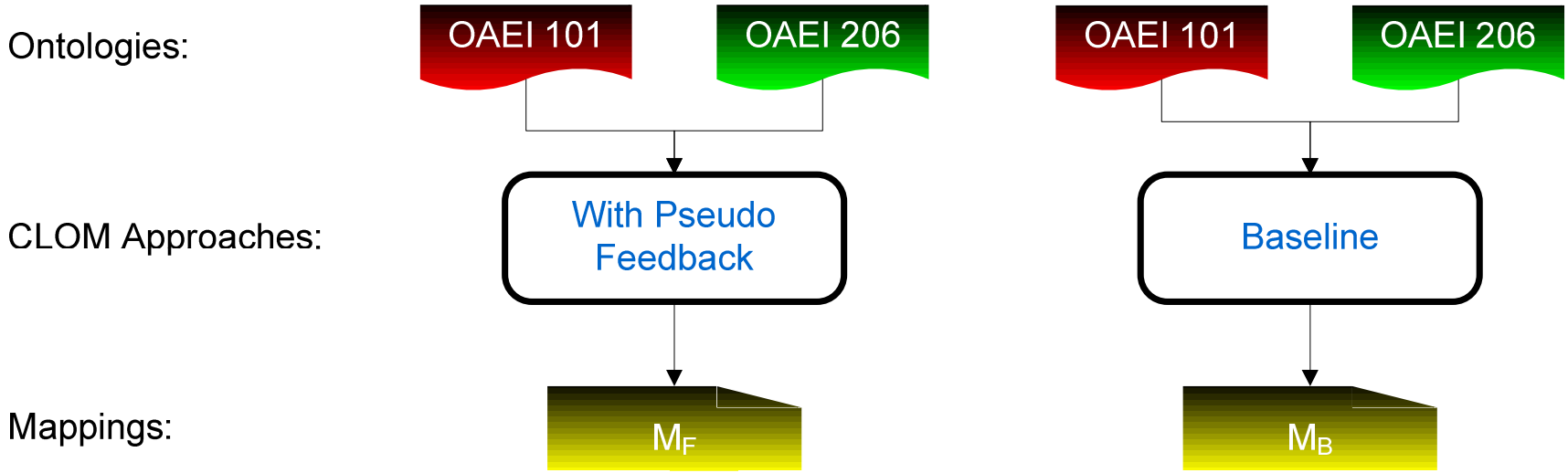
- Generate initial set of matches
- Make assumptions on their correctness
- Compute a better set of matches based on this assumption


O1
O2
jena
Machine Translation Service
Windows Live
Google translate
Source Labels
Target Labels
Lexicon Dictionary Service
eXist
Translation Repository
Source Structure
Target Structure
eXist
Lexicon Repository
AOLT Selection
Interpretative Keyword Generation Service
YAHOO!
WIKIPEDIA
The Free Encyclopedia
O1 AOLTs
jena
O1'
Pseudo Feedback for Run 2
Monolingual Ontology Matching Service
INRIA
Candidate Matches
Match Assessment
Mappings

Collision Resolution

Collision Scenario	AOLT Selection Strategy		Solution
	E_1	E_2	
i	Candidate translation = target label's synonym	Candidate translation = target label	E_2 keeps collided term; E_1 seeks alternative AOLT
ii	Derived from interpretative keyword comparison	Candidate translation = target label's synonym	
iii	Candidate translation = target label	Candidate translation = target label's synonym	E_1 keeps collided term; E_2 seeks alternative AOLT
iv	Candidate translation = target label's synonym	Derived from interpretative keyword comparison	
v	Candidate translation = target label	Candidate translation = target label	
vi	Candidate translation = target label's synonym	Candidate translation = target label's synonym	
vii	Derived from interpretative keyword comparison	Derived from interpretative keyword comparison	

Evaluation

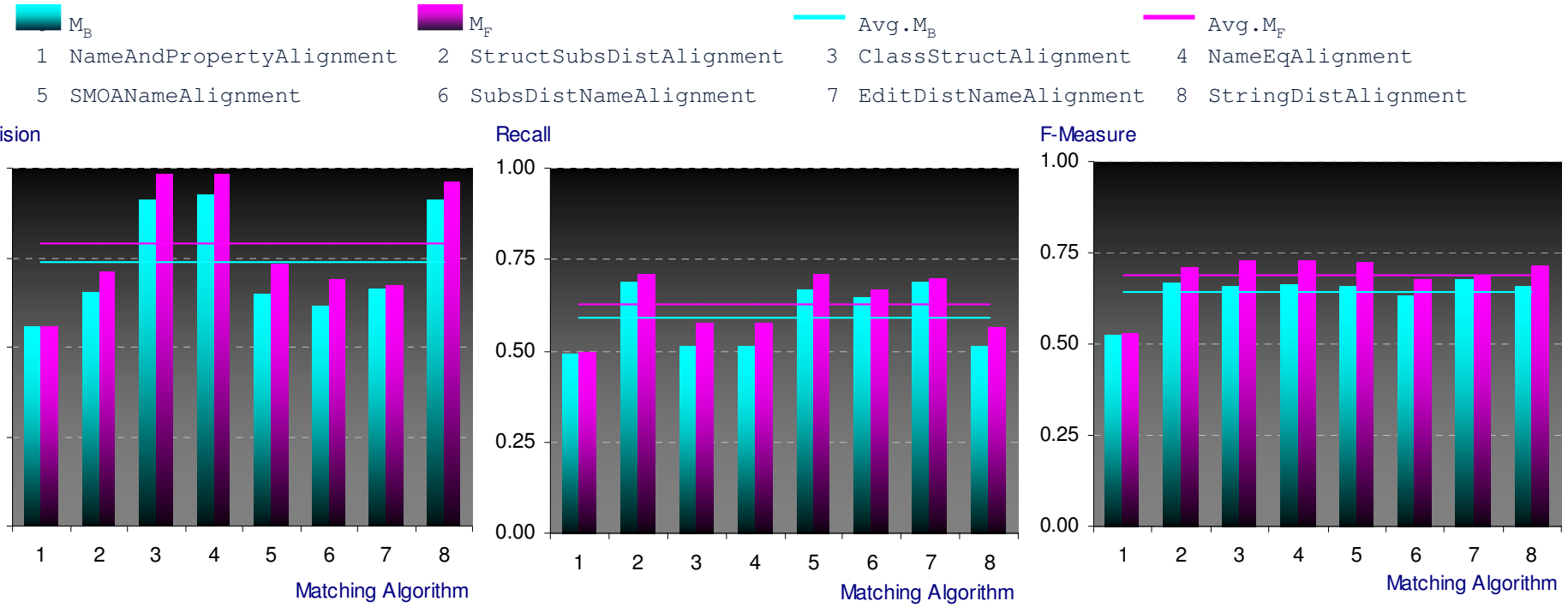


Matching Method **Assumption** **Total Matches** **"Correct" Matches**

```
<PseudoFeedback algorithm="StrucSubsDistAlignment" threshold="0.5" matches="103.0" estimate="86.0" >
  <Entry count="31.0" medium="TargetOntology" usage="0.360465"/>
  <Entry count="23.0" medium="BothMT" usage="0.267441"/>
  <Entry count="17.0" medium="Google" usage="0.197674"/>
  <Entry count="12.0" medium="WindowsLive" usage="0.139534"/>
  <Entry count="3.0" medium="External" usage="0.034883"/>
</PseudoFeedback>
```

A Ranked List of Translation Media

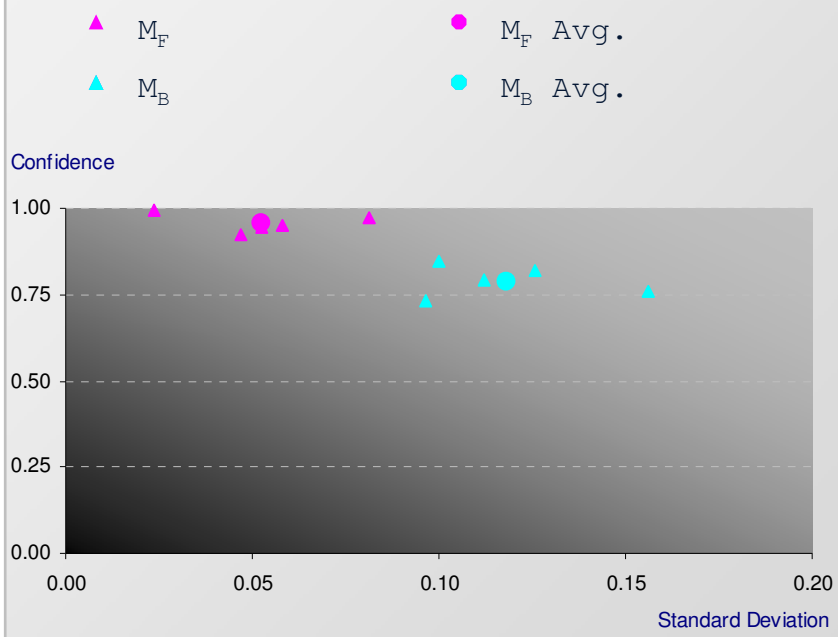
Results



	Precision	Recall	F-Measure
Avg. M_B	.7355	.5928	.6428
Avg. M_F	.7875	.6268	.6873
Improvement (%)	7.07	5.74	6.92

Results

Confidence Mean & St. Dev.



Two-Tailed Paired T-Test on F-Measure

Null Hypothesis:

H_0: $M_B = M_F$ (there is no difference between M_B and M_F);

Alternative Hypothesis:

H_a: $M_B \neq M_F$ (there is a difference between M_B and M_F);

P-Value = 0.003 rejects null hypothesis at a significance level of $\alpha = 0.05$

Conclusions

- Motivating findings from the pseudo feedback mechanism
- Limitations & Future Work
  - Feedback
    - Explicit
    - Implicit
    - User-configured cut-off points
    - Negative feedback
  - Experiments
    - Specialised domain and MT tools
    - Distinct natural language pairs
    - Scalability test with large ontology sets
  - Evaluation
    - Impact on mapping quality from invalid pseudo feedback assumptions
    - Third, fourth, fifth etc. iterations of the feedback mechanism

Acknowledgement

This research is partially supported by Science Foundation Ireland (Grant 07/CE/11142) as part of the Centre for Next Generation Localisation (www.cngl.ie), and the Higher Education Authority (PRTLI4 fund) as part of Networked Embedded Systems project (www.nembes.org) at Trinity College Dublin.

Thank You!