# Overview of BioNLP'09 Shared Task on Event Extraction

Jin-Dong Kim, Tomoko Ohta, Sampo Pyysalo,
Yoshinobu Kano, Jun'ichi Tsujii

University of Tokyo

# Contents

- Introduction
- Task Setting
- Data Preparation
- Schedule
- Participation
- Evaluation methods
- Results
- Conclusion
- Discussion

# Shared Tasks in BioTM community

- Shared Tasks in Bio-TextMining
  - ✓ TREC Genomics track (2003~2007) IR
  - ✓ BioNLP (2004) NER
  - ✓ LLL (2005) IE
  - ✓ BioCreative 2005, 2007 IE
    - ➲ Focuses on PPI (unspecified relation between two proteins)
    - ➲ Goal: to support curation of PPI databases (MINT)
    - ➲ Extrinsic evaluation - actual contribution to PPI curation task.
  - ✓ BioNLP 2009 Event Extraction IE
    - ➲ Focuses on Events (detailed behavior of proteins, Ontology)
    - ➲ Goal: to provide resources for improvement.
    - ➲ Intrinsic evaluation - general event extraction performance
      - – Final application is not fixed.
      - – The results should be **interpretable** to seek **further improvement**.

# Contents

- Introduction
- Task Setting
- Data Preparation
- Schedule
- Participation
- Evaluation methods
- Results
- Conclusion
- Discussion

# Target Event Types

| Event Type | |
|---|---|
| Gene_expression | Protein Metabolism |
| Transcription | |
| Protein_catabolism | |
| Phosphorylation | |
| Localization | Fundamental Molecular Events |
| Binding | |
| Regulation | Regulatory events/relation |
| Positive_regulation | |
| Negative_regulation | |

- We considered
  - ✓ Biological implication
  - ✓ Complexity as computational tasks
  - ✓ Number and quality of annotation instances (in the GENIA corpus)

# Target Event Types

| Event Type | Primary Arguments |
|---|---|
| Gene_expression | Theme(Prot) |
| Transcription | Theme(Prot) |
| Protein_catabolism | Theme(Prot) |
| Phosphorylation | Theme(Prot) |
| Localization | Theme(Prot) |
| Binding | Theme(Prot)+ |
| Regulation | Theme(Prot/Ev), Cause (Prot/Ev) |
| Positive_regulation | Theme(Prot/Ev), Cause (Prot/Ev) |
| Negative_regulation | Theme(Prot/Ev), Cause (Prot/Ev) |

Simple: 1 argument

Complex: n arguments

Complex: 2 arguments, recursive

- We considered
  - ✓ Biological implication
  - ✓ Complexity as computational tasks
  - ✓ Number and quality of annotation instances (in the GENIA corpus)

# Target Event Types

| Event Type | Primary Arguments | Secondary Arguments |
|---|---|---|
| Gene_expression | Theme(Prot) | |
| Transcription | Theme(Prot) | |
| Protein_catabolism | Theme(Prot) | |
| Phosphorylation | Theme(Prot) | Site(Entity) |
| Localization | Theme(Prot) | AtLoc(Entity), ToLoc(Entity) |
| Binding | Theme(Prot)+ | Site(Entity)+ |
| Regulation | Theme(Prot/Ev), Cause (Prot/Ev) | Site(Entity), CSite(Entity) |
| Positive_regulation | Theme(Prot/Ev), Cause (Prot/Ev) | Site(Entity), CSite(Entity) |
| Negative_regulation | Theme(Prot/Ev), Cause (Prot/Ev) | Site(Entity), CSite(Entity) |

- We considered
  - ✓ Biological implication
  - ✓ Complexity as computational tasks
  - ✓ Number and quality of annotation instances (in the GENIA corpus)

# Task Definition

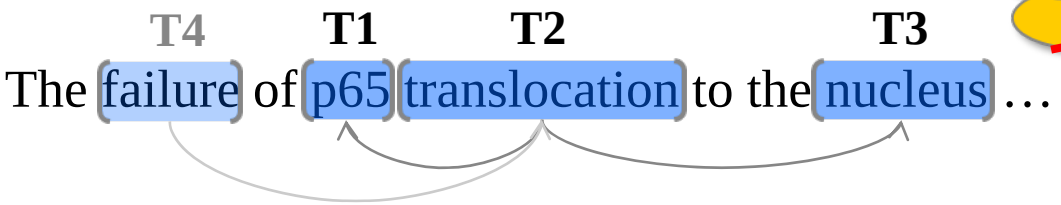
- Tasks
  - ✓ Task 0: Protein recognition (Given)
  - ✓ Task 1: Core event extraction
  - ✓ Task 2: Event enrichment
  - ✓ Task 3: Negation/Speculation detection

- The only feature that makes it less real.
- To draw concentration on event extraction.
- Makes evaluation results better interpretable.

non-text-bound IE

~~T4 (Negation, 4-11)~~
~~M1 (Mod:T4 Event:E1)~~

- Example




| | | |
|---|---|---|
| T1 | (Protein, 15-18) | Protein recognition (GIVEN) |
| T2 | (Localization, 19-32) | Trigger recognition (Task 1) |
| T3 | (Entity, 40-46) | Entity recognition (Task 2) |
| E1 | (Type:T2, Theme:T1, ToLoc:T3) | Argument detection (Task 1, 2) |
| M1 | (Negation Event:E1) | Negation/Speculation (Task 3) |

# Contents

- Introduction
- Task Setting
- Data Preparation
- Schedule
- Participation
- Evaluation methods
- Results
- Conclusion
- Discussion

# Data Preparation

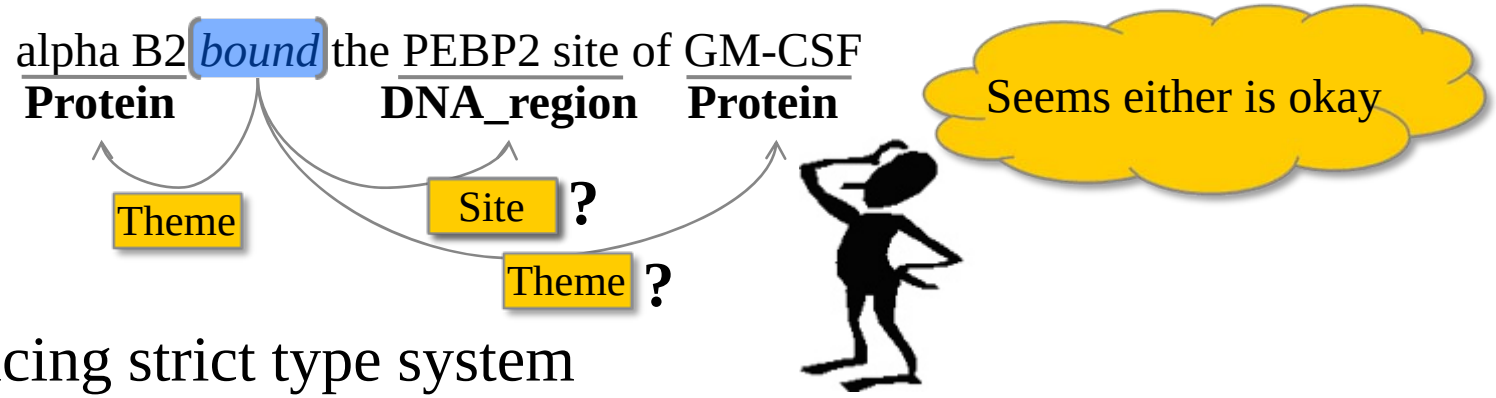
- Shared task data sets were produced
  - ✓ from the GENIA event corpus
    - ➲ publicly available 1,000 abstracts → training & development sets
    - ➲ internally available 1,000 abstracts → test set
- Processes for data production
  - ✓ Filtering for only entities and events relevant to target event types.
  - ✓ gene-or-gene-product annotation.
  - ✓ Argument revision.
  - ✓ Equivalent entity annotation.

# Gene-or-gene-product annotation

- GENIA entity annotation
  - ✓ Differentiates protein, gene (DNA region) and RNA.
- GENETAG (BioCreative), AIMed, LLL corpora
  - ✓ Does not differentiate protein, gene (DNA region) and RNA
    - ➲ Protein (gene)
- Definitions
  - ✓ **Gene** XXX encodes how to implement *a function*.
  - ✓ **RNA** XXX mediates the information (how to implement *the function*).
  - ✓ **protein** XXX performs *the function*.
- GGP annotation to GENIA
  - ✓ Does not differentiate protein, gene (DNA region) and RNA.
  - ✓ Comparable to protein (or gene) annotation of other corpora.
- For more details,
  - ✓ "*Incorporating GENETAG-style annotation to GENIA corpus*", Tomoko Ohta, Jin-Dong Kim, Sampo Pyysalo, Yue Wang and Jun'ichi Tsujii, BioNLP 2009 NAACL-HLT Workshop.

# Argument revision

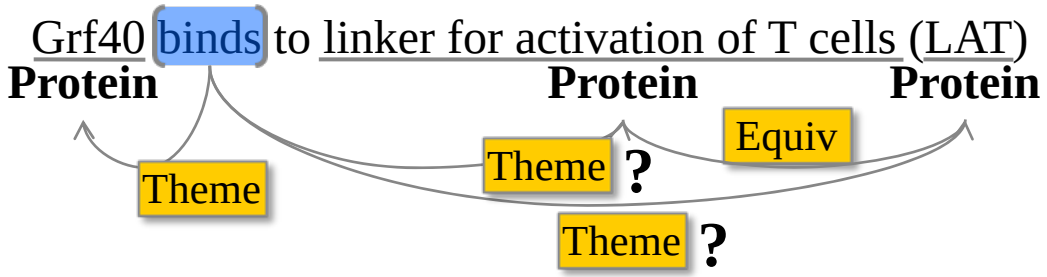
- GENIA event annotation
  - ✓ was created based on loose type system
    - ➲ ex) Theme of Binding event: bio-entity




- Introducing strict type system
  - ✓ ex) Theme of Binding event: Protein
  - ✓ For it, annotation for static relations was added.
- For more detail,
  - ✓ *Static Relations: a Piece in the Biomedical Information Extraction Puzzle* Sampo Pyysalo, Tomoko Ohta, Jin-Dong Kim and Jun'ichi Tsujii, BioNLP 2009 NAACL-HLT Workshop

# Equivalent Entities

- GENIA event corpus
  - ✓ When there are equivalent entity references, only one of them is chosen to be an argument of an event.




- Equivalent entity annotation
  - ✓ Candidates collection
    - ➲ Patterns collected from GENIA coreference corpus, manually revised.
      - – [PROT (PROT)], [PROT/PROT], [PROT, PROT], [PROT: PROT], [PROT, namely PROT], [PROT, also called PROT], …
  - ✓ Manual decision
    - ➲ 1 biologist

# Statistics of Data Sets

| Type | Training | | Devel. | | Test | |
|---|---|---|---|---|---|---|
| ABSTRACT | 800 | | 150 | | 260 | |
| | T1 | T2 | T1 | T2 | T1 | T2 |
| Gene_exp. | 1738 | | 356 | | 722 | |
| Transcription | 576 | | 81 | | 137 | |
| Prot._catabolism | 110 | | 21 | | 14 | |
| Phosphorylation | 165 | 169 | 47 | 47 | 135 | 139 |
| Localization | 263 | 265 | 53 | 53 | 174 | 174 |
| Binding | 880 | 887 | 248 | 249 | 347 | 349 |
| Regulation | 960 | 961 | 174 | 178 | 291 | 292 |
| Posi._regulation | 2843 | 2847 | 632 | 633 | 983 | 987 |
| Nega._regulation | 1062 | 1062 | 197 | 197 | 379 | 379 |

# Schedule

- ❑ Announcement (Dec 8)
- ❑ Sample Data (Dec 15)
- ❑ Training Data (Jan 19 → 21)
  - ✓ rev 1 (Feb 2)
  - ✓ rev 2 (Feb 10)
- ❑ Devel. Data (Feb 7)
  - ✓ with online evaluation system
- ❑ Test Data (Feb 22 → Mar 2)
- ❑ Submission (Mar 2 → Mar 9)

System Design period (5 weeks)

System Devel. Period (5 weeks → 6 weeks)

Test period (1 week)

# Supporting Resources

- Purpose
  - ✓ To allow participants to concentrate on new aspects of event extraction.
- Easy access to publicly available NLP tools was provided.
  - ✓ Fundamental NLP tools through U-Compare
    - ➲ Tokenization
    - ➲ sentence segmentation
    - ➲ part-of-speech tagging
    - ➲ Chunking
    - ➲ parsing
  - ✓ Syntactically parsed data sets, by
    - ➲ Dan Bikel's implementation of the Collins' parsing model. ("Bikel")
    - ➲ The Charniak-Johnson reranking parser using David McClosky's self-trained biomedical parsing model. ("McClosky-Charniak")
    - ➲ GDep, a version of the LRDEP/KSDEP native dependency parser trained on the GENIA Treebank
    - ➲ C&C CCG parser

# Participation

- 42 teams registered
- 24 teams submitted final results
  - ✓ 24 teams for Task 1
  - ✓ 6 teams for Task 2
  - ✓ 6 teams for Task 3
  - ✓ (2 teams for the all three tasks)

# Team profiles

| Team | Task | Org | NLP | | | Task | | Ext. Resources |
|---|---|---|---|---|---|---|---|---|
| | | | Word | Chunking | Parsing | Trigger | Argument | |
| UTurku | 1-- | 3C+2BI | Porter | | MC | SVM | SVM (SVMlight) | |
| JULIELab | 1-- | 1C+2L+2B | OpenNLP<br>Porter | OpenNLP | GDep | Dict+Stat | SVM(libSVM)<br>ME(Mallet) | UniProt, Mesh, GOA, UMLS |
| ConcordU | 1-3 | 3C | Stanford | | Stanford | Dict+Stat | Rules | WordNet, VerbNet, UMLS |
| UT+DBCLS | 12- | 2C | Porter | | MC<br>CCG | Dict | MLN(thebeast) | |
| VIBGhent | 1-3 | 2C+1B | Porter, | | Stanford | Dict | SVM(libSVM) | |
| UTokyo | 1-- | 3C | GTag | | GDep,<br>Enju | Dict | ME(liblinear) | UIMA |
| UNSW | 1-- | 1C+1B | | | GDep | CRF | Rules | WordNet, MetaMap |
| UZurich | 1-- | 3C | LingPipe,<br>Morpha | LTChunk | Pro3Gres | Dict | Rules | |
| ASU+HU+BU | 123 | 6C+2BI | Porter | | BioLG,<br>Charniak | Dict | Rules<br>Rules | Lucene |
| Cam | 1-- | 3C | Porter | | RASP | Dict | Rules | |
| UAntwerp | 12- | 3C | GTag | | GDep | MBL | MBL(TiMBL)<br>Rules | |
| UNIMAN | 1-- | 4C+2BI | Porter<br>GTag | | GDep | Dict, CRF | SVM<br>Rules | MeSH, GO |
| SCAI | 1-- | 1C | | | | | Rules | |
| UAveiro | 1-- | 1C+1L | NooJ | NooJ | | | Rules | BioLexicon |
| USzeged | 1-3 | 3C+1B | GTag | | | Dict, VSM | C4.5(WEKA)<br>Rules | BioScope |
| NICTA | 1-3 | 4C | GTag | | ERG | CRF(CRF++) | Rules | JULIE |
| CNBMadrid | 12- | 2C+1B | Porter,<br>GTag | GTag | | | CBR<br>Rules | |
| CCP-BTMG | 123 | 7C | LingPipe | LingPipe | OpenDMAP | LingPipe, CM | Rules | GO, SO, MIO, UIMA |
| CIPS-ASU | 1-- | 3C | MontyTagger | Custom | Stanford | CRF(ABNER) | Rules,<br>NB(WEKA) | |
| UMich | 1-- | 2C | Stanford | | MC | Dict | SVM(SVMlight) | |
| PIKB | 1-- | 5C+2B | | | | MIRA | MIRA | |
| KoreaU | 1-- | 5C | GTag | | GDep | Rules, ME | ME | WSJ |

Table 4: Profiles of the participants: GTag=GENIAtagger, MLN=Markov Logic Network, UMLS=UMLS SPECIALIST Lexicon/tools, MC=McClosky-Charniak, GDep=Genia Dependency Parser, Stanford=Stanford Parser, CBR=Case-Based Reasoning, CM=ConceptMapper.

# Team profiles

| Team | Task | Org | NLP: Word | NLP: Chunking | NLP: Parsing | Task: Trigger | Task: Argument | Ext. Resources |
|---|---|---|---|---|---|---|---|---|
| UTurku | 1-- | 3C+2BI | | | | | SVMlight) | |
| JULIELab | 1-- | 1C+2L+2B | | | | | (libSVM) (Mallet) | UniProt, Mesh, GOA, UMLS |
| ConcordU | 1-3 | 3C | | | | | Rules | WordNet, VerbNet, UMLS |
| UT+DBCLS | 12- | 2C+1B | | | | | (thebeast) | |
| VIBGhent | 1-3 | 3C | | | | | (libSVM) | |
| UTokyo | 1-- | 1C+1B | | | | | (liblinear) | UIMA |
| UNSW | 1-- | 3C | | | | | Rules | WordNet, MetaMap |
| UZurich | 1-- | 6C+2BI | LingPipe, | LTChunk | Pro3Gres | Dict | Rules | |
| ASU+HU+BU | 123 | 3C | | | | | | Lucene |
| Cam | 1-- | 3C | | | | | | |
| UAntwerp | 12- | | | | | | | |
| UNIMAN | 1-- | 4C+2BI | | | | | | MeSH, GO |
| SCAI | 1-- | 1C | | | | | | |
| UAveiro | 1-- | 1C+1L | | | | | | BioLexicon |
| USzeged | 1-3 | 3C+1B | | | | | | BioScope |
| NICTA | 1-3 | 4C | | | | | | JULIE |
| CNBMadrid | 12- | 2C+1B | | | | | | |
| CCP-BTMG | 123 | 7C | LingPipe | LingPipe | OpenDMAP | LingPipe, CM | Rules | GO, SO, MIO, UIMA |
| CIPS-ASU | 1-- | 3C | MontyTagger | Custom | Stanford | CRF(ABNER) | Rules, NB(WEKA) | |
| UMich | 1-- | 2C | Stanford | | MC | Dict | SVM(SVMlight) | |
| PIKB | 1-- | 5C+2B | | | | MIRA | MIRA | |
| KoreaU | 1-- | 5C | GTag | | GDep | Rules, ME | ME | WSJ |

67 Computer scientists (C)
8 Biologists (B)
6 Bioinformaticians (BI)
4 Linguists (L)

New and complex task

→Computational modeling is important

Once computational model gets stable input from biologists will become important.

Table 4: Profiles c ants: GTag=GENIAtagger, MLN=Markov Logic Network, UMLS=UMLS SPECIALIST Lexicon/ cClosky-Charniak, GDep=Genia Dependency Parser, Stanford=Stanford Parser, CBR=Case-Based I =ConceptMapper.

# Team profiles

| Team | Task | Org | NLP | | | Task | | Ext. Resources |
|---|---|---|---|---|---|---|---|---|
| | | | Word | Chunking | Parsing | Trigger | Argument | |
| UTurku | 1-- | 3C+2BI | Porter | | MC | SVM | SVM (SVMlight) | |
| JULIELab | 1-- | 1C+2L+2B | OpenNLP<br>Porter | OpenNLP | GDep | Dict+Stat | SVM(libSVM)<br>ME(Mallet) | UniProt, Mesh, GOA, UMLS |
| ConcordU | 1-3 | 3C | Stanford | | Stanford | Dict+Stat | Rules | WordNet, VerbNet, UMLS |
| UT+DBCLS | 12- | 2C | Porter | | MC<br>CCG | Dict | MLN(thebeast) | |
| VIBGhent | 1-3 | 2C+1B | Porter, | | Stanford | Dict | SVM(libSVM) | |
| UTokyo | 1-- | 3C | GTag | | GDep,<br>Enju | Dict | ME(liblinear) | UIMA |
| UNSW | 1-- | 1C+1B | | | GDep | CRF | Rules | WordNet, MetaMap |
| UZurich | 1-- | 3C | LingPipe,<br>Morpha | LTChunk | Pro3Gres | Dict | Rules | |
| ASU+HU+BU | 123 | 6C+2BI | Porter | | BioLG,<br>Charniak | Dict | Rules<br>Rules | Lucene |
| Cam | 1-- | 3C | Porter | | RASP | Dict | Rules | |
| UAntwerp | 12- | 3C | GTag | | GDep | MBL | MBL(TiMBL)<br>Rules | |
| UNIMAN | 1-- | 4C+2BI | Porter<br>GTag | | GDep | Dict, CRF | SVM<br>Rules | MeSH, GO |
| SCAI | 1-- | 1C | | | | | Rules | |
| UAveiro | 1-- | 1C+1L | NooJ | NooJ | | | Rules | BioLexicon |
| USzeged | 1-3 | 3C+1B | GTag | | | Dict, VSM | C4.5(WEKA)<br>Rules | BioScope |
| NICTA | 1-3 | 4C | GTag | | ERG | CRF(CRF++) | Rules | JULIE |
| CNBMadrid | 12- | 2C+1B | Porter,<br>GTag | GTag | | | CBR<br>Rules | |
| CCP-BTMG | 123 | 7C | LingPipe | LingPipe | OpenDMAP | LingPipe, CM | Rules | GO, SO, MIO, UIMA |
| CIPS-ASU | 1-- | 3C | MontyTagger | Custom | Stanford | CRF(ABNER) | Rules,<br>NB(WEKA) | |
| UMich | 1-- | 2C | Stanford | | MC | Dict | SVM(SVMlight) | |
| PIKB | 1-- | 5C+2B | | | | MIRA | MIRA | |
| KoreaU | 1-- | 5C | GTag | | GDep | Rules, ME | ME | WSJ |

Table 4: Profiles of the participants: GTag=GENIAtagger, MLN=Markov Logic Network, UMLS=UMLS SPECIALIST Lexicon/tools, MC=McClosky-Charniak, GDep=Genia Dependency Parser, Stanford=Stanford Parser, CBR=Case-Based Reasoning, CM=ConceptMapper.

# Full parsing

**- Applied by 73% of the all systems**

**- Applied by the all systems in top 50%**

**Strong support for the value of full parsing?**

# Te ofiles

| | | | | Parsing | Task | | |
|---|---|---|---|---|---|---|---|
| | | | | | Trigger | Argument | Ext. Resources |
| | | | | Stanford | SVM | SVM (SVMlight) | |
| | | | | MC CCG | Dict+Stat | SVM(libSVM) ME(Mallet) | UniProt, Mesh, GOA, UMLS |
| | | | | Stanford | Dict+Stat | Rules | WordNet, VerbNet, UMLS |
| | | | | GDep, Enju | Dict | MLN(thebeast) | |
| | | | | GDep | Dict | SVM(libSVM) | |
| | | | | Pro3Gres | Dict | ME(liblinear) | UIMA |
| | | | | | CRF | Rules | WordNet, MetaMap |
| | | | | BioLG, Charniak | Dict | Rules | |
| | | | | RASP | Dict | Rules Rules | Lucene |
| | | | | GDep | Dict | Rules | |
| | | | | | MBL | MBL(TiMBL) Rules | |
| | | | | GDep | Dict, CRF | SVM Rules | MeSH, GO |
| | | | | | | Rules | |
| | | | | | | Rules | BioLexicon |
| USzeged | 1-3 | 3C+1B | GTag | | Dict, VSM | C4.5(WEKA) Rules | BioScope |
| NICTA | 1-3 | 4C | GTag | | CRF(CRF++) | Rules | JULIE |
| CNBMadrid | 12- | 2C+1B | Porter, GTag | GTa | | CBR Rules | |
| | | | | ERG | | | |
| CCP-BTMG | 123 | 7C | LingPipe | LingP | LingPipe, CM | Rules | GO, SO, MIO, UIMA |
| CIPS-ASU | 1-- | 3C | MontyTagger | Cust | CRF(ABNER) | Rules, NB(WEKA) | |
| | | | | OpenDMAP | | | |
| UMich | 1-- | 2C | Stanford | | Dict | SVM(SVMlight) | |
| PIKB | 1-- | 5C+2B | | | MIRA | MIRA | |
| KoreaU | 1-- | 5C | GTag | Stanford | Rules, ME | ME | WSJ |
| | | | | MC | | | |
| | | | | GDep | | | |

Table 4: Profiles of the participants: GTag=GE N=Markov Logic Network, UMLS=UMLS SPECIALIST Lexicon/tools, MC=McClosky-Charnia a Dependency Parser, Stanford=Stanford Parser, CBR=Case-Based Reasoning, CM=ConceptMappe

# Team profiles

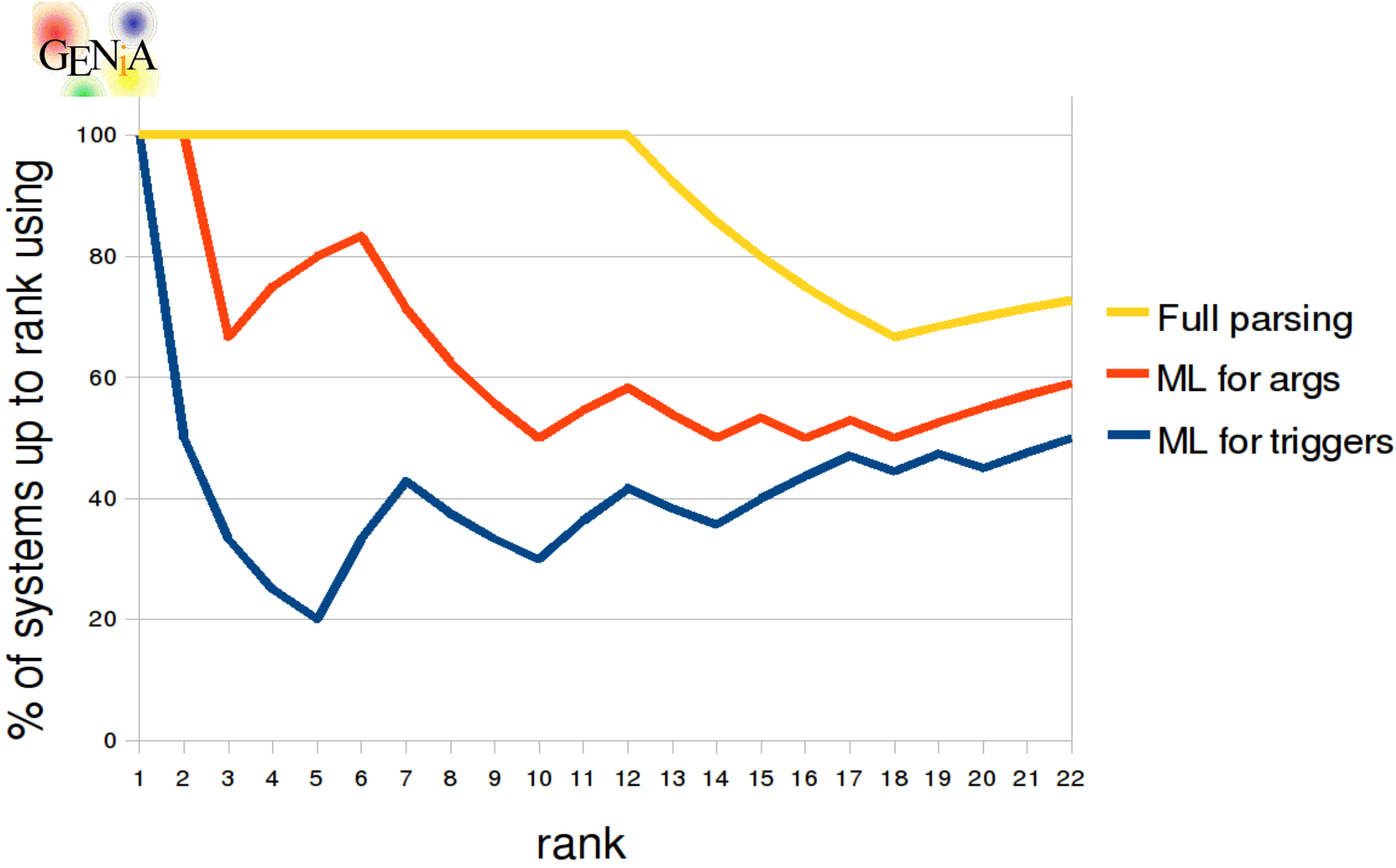
| Team | Task | Org | NLP | | | Task | | Ext. Resources |
|---|---|---|---|---|---|---|---|---|
| | | | Word | Chunking | Parsing | Trigger | Argument | |
| UTurku | 1-- | 3C+2BI | Porter | | MC | SVM | SVM (SVMlight) | |
| JULIELab | 1-- | 1C+2L+2B | OpenNLP<br>Porter | OpenNLP | GDep | Dict+Stat | SVM(libSVM)<br>ME(Mallet) | UniProt, Mesh, GOA, UMLS |
| ConcordU | 1-3 | 3C | Stanford | | Stanford | Dict+Stat | Rules | WordNet, VerbNet, UMLS |
| UT+DBCLS | 12- | 2C | Porter | | MC<br>CCG | Dict | MLN(thebeast) | |
| VIBGhent | 1-3 | 2C+1B | Porter, | | Stanford | Dict | SVM(libSVM) | |
| UTokyo | 1-- | 3C | GTag | | GDep,<br>Enju | Dict | ME(liblinear) | UIMA |
| UNSW | 1-- | 1C+1B | | | GDep | CRF | Rules | WordNet, MetaMap |
| UZurich | 1-- | 3C | LingPipe,<br>Morpha | LTChunk | Pro3Gres | Dict | Rules | |
| ASU+HU+BU | 123 | 6C+2BI | Porter | | BioLG,<br>Charniak | Dict | Rules<br>Rules | Lucene |
| Cam | 1-- | 3C | Porter | | RASP | Dict | Rules | |
| UAntwerp | 12- | 3C | GTag | | GDep | MBL | MBL(TiMBL)<br>Rules | |
| UNIMAN | 1-- | 4C+2BI | Porter<br>GTag | | GDep | Dict, CRF | SVM<br>Rules | MeSH, GO |
| SCAI | 1-- | 1C | | | | | Rules | |
| UAveiro | 1-- | 1C+1L | NooJ | NooJ | | | Rules | BioLexicon |
| USzeged | 1-3 | 3C+1B | GTag | | | Dict, VSM | C4.5(WEKA)<br>Rules | BioScope |
| NICTA | 1-3 | 4C | GTag | | ERG | CRF(CRF++) | Rules | JULIE |
| CNBMadrid | 12- | 2C+1B | Porter,<br>GTag | GTag | | | CBR<br>Rules | |
| CCP-BTMG | 123 | 7C | LingPipe | LingPipe | OpenDMAP | LingPipe, CM | Rules | GO, SO, MIO, UIMA |
| CIPS-ASU | 1-- | 3C | MontyTagger | Custom | Stanford | CRF(ABNER) | Rules,<br>NB(WEKA) | |
| UMich | 1-- | 2C | Stanford | | MC | Dict | SVM(SVMlight) | |
| PIKB | 1-- | 5C+2B | | | | MIRA | MIRA | |
| KoreaU | 1-- | 5C | GTag | | GDep | Rules, ME | ME | WSJ |

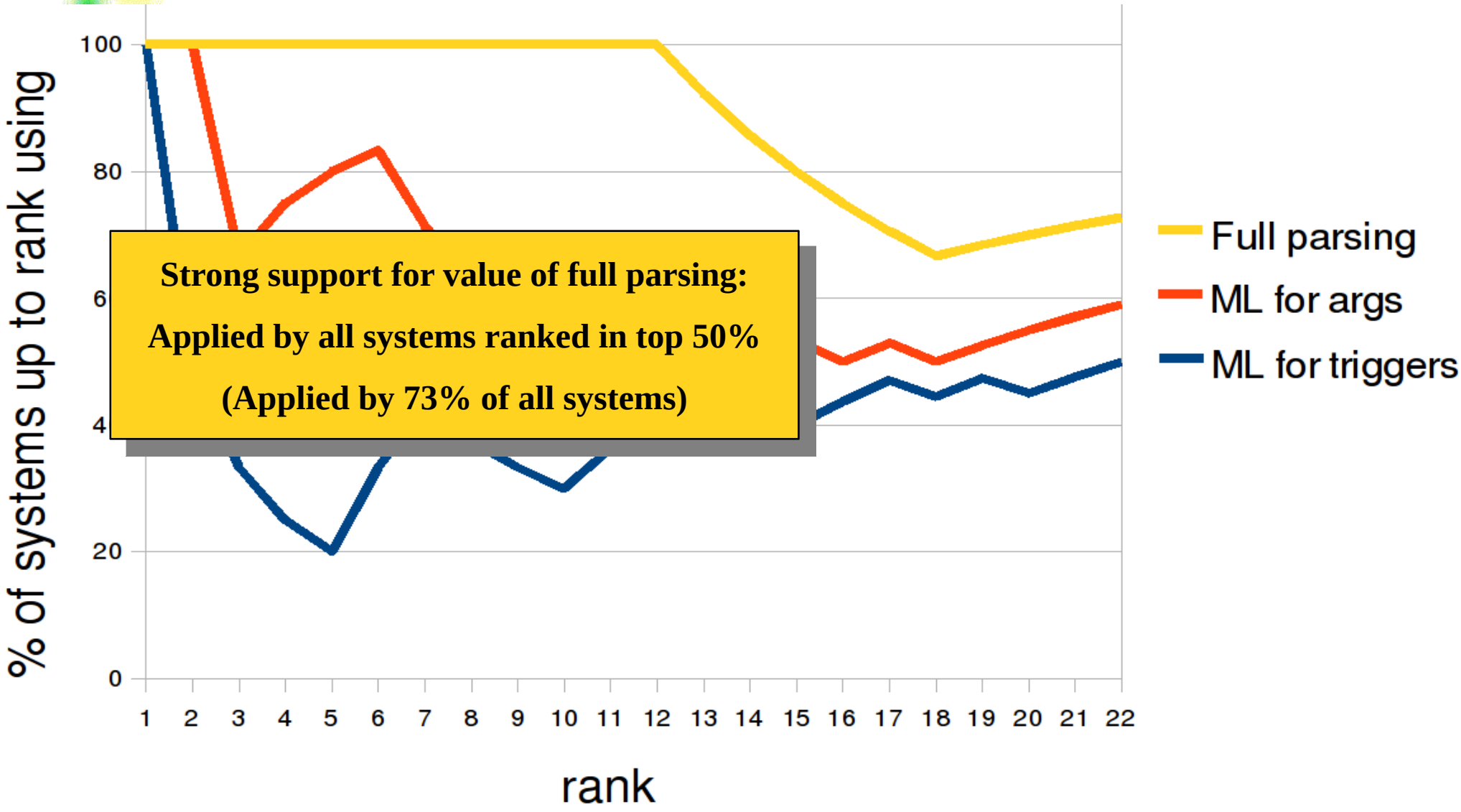
Table 4: Profiles of the participants: GTag=GENIAtagger, MLN=Markov Logic Network, UMLS=UMLS SPECIALIST Lexicon/tools, MC=McClosky-Charniak, GDep=Genia Dependency Parser, Stanford=Stanford Parser, CBR=Case-Based Reasoning, CM=ConceptMapper.

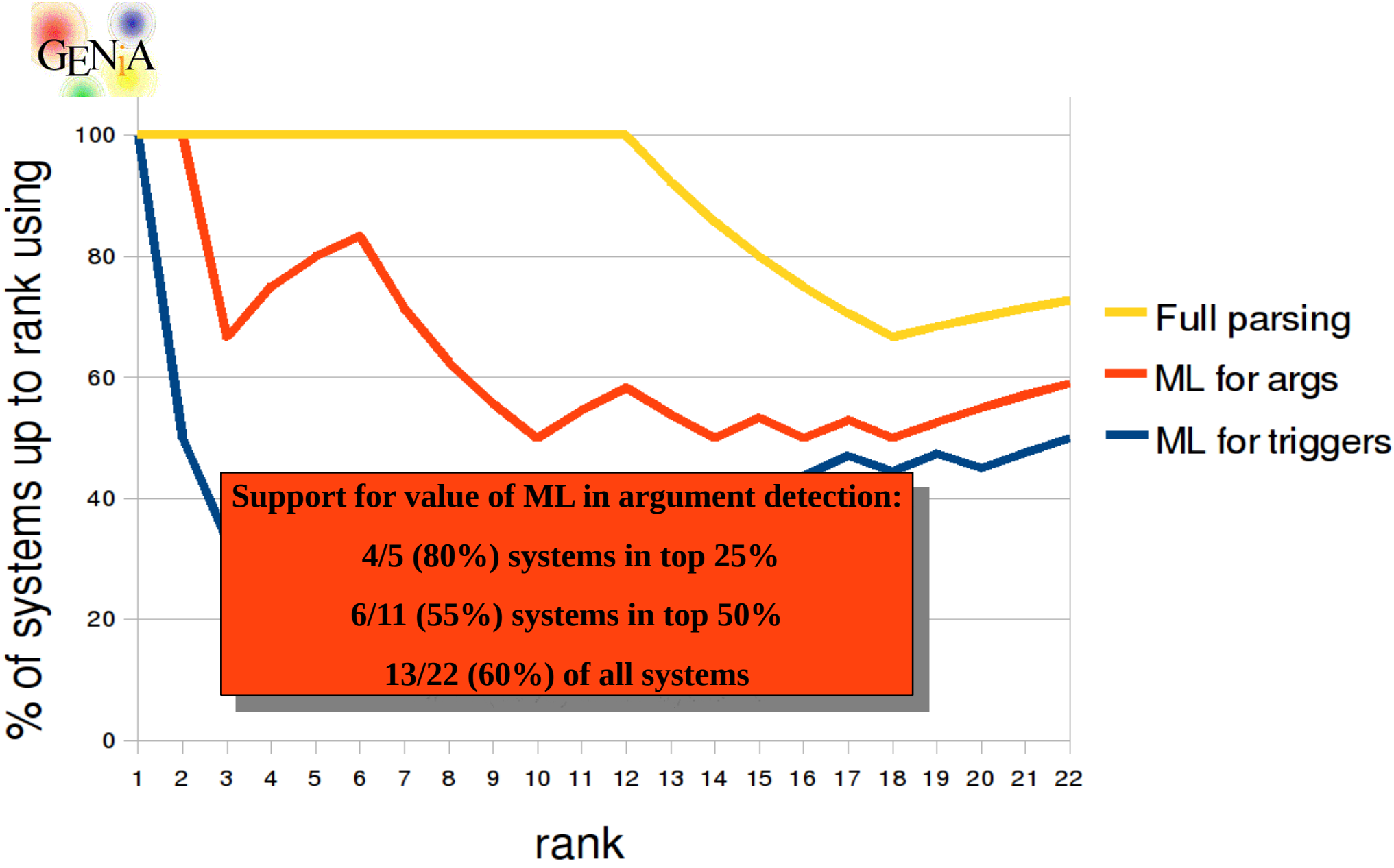
# Team profil

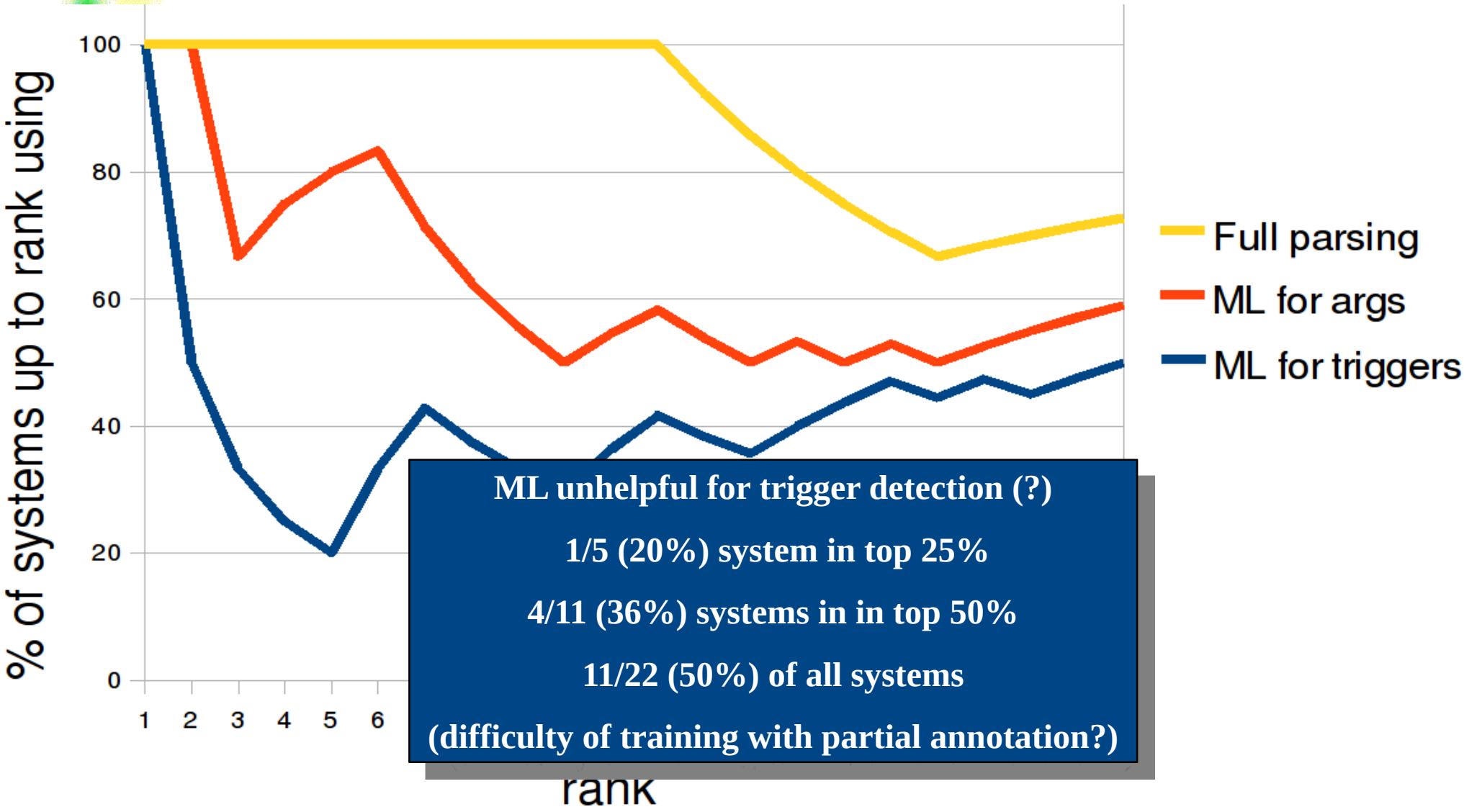
| Team | Task | Org | NLP | | | Trig | Argument | Ext. Resources |
|---|---|---|---|---|---|---|---|---|
| | | | Word | Chunking | Parsing | | | |
| UTurku | 1-- | 3C+2BI | Porter | | MC | SV | SVM (SVMlight) | |
| JULIELab | 1-- | 1C+2L+2B | OpenNLP Porter | OpenNLP | GDep | Dict+ | SVM(libSVM) ME(Mallet) | UniProt, Mesh, GOA, UMLS |
| ConcordU | 1-3 | 3C | Stanford | | Stanford | Dict+ | Rules | WordNet, VerbNet, UMLS |
| UT+DBCLS | 12- | 2C | Porter | | MC CCG | Di | MLN(thebeast) | |
| VIBGhent | 1-3 | 2C+1B | Porter, | | Stanford | Di | SVM(libSVM) | |
| UTokyo | 1-- | 3C | GTag | | GDep, Enju | Di | ME(liblinear) | UIMA |
| UNSW | 1-- | 1C+1B | | | GDep | CR | Rules | WordNet, MetaMap |
| UZurich | 1-- | 3C | LingPipe, Morpha | LTChunk | Pro3Gres | Di | Rules | |
| ASU+HU+BU | 123 | 6C+2BI | Porter | | BioLG, Charniak | Di | Rules Rules | Lucene |
| Cam | 1-- | 3C | Porter | | RASP | Di | Rules | |
| UAntwerp | 12- | 3C | GTag | | GDep | ME | MBL(TiMBL) Rules | |
| UNIMAN | 1-- | 4C+2BI | Porter GTag | | GDep | Dict, | SVM Rules | MeSH, GO |
| SCAI | 1-- | 1C | | | | | Rules | |
| UAveiro | 1-- | 1C+1L | NooJ | NooJ | | | Rules | BioLexicon |
| USzeged | 1-3 | 3C+1B | GTag | | | Dict, | C4.5(WEKA) Rules | BioScope |
| NICTA | 1-3 | 4C | GTag | | ERG | CRF(C | Rules | JULIE |
| CNBMadrid | 12- | 2C+1B | Porter, GTag | GTag | | | CBR Rules | |
| CCP-BTMG | 123 | 7C | LingPipe | LingPipe | OpenDMAP | LingPip | Rules | GO, SO, MIO, UIMA |
| CIPS-ASU | 1-- | 3C | MontyTagger | Custom | Stanford | CRF(AI | Rules, NB(WEKA) | |
| UMich | 1-- | 2C | Stanford | | MC | Di | SVM(SVMlight) | |
| PIKB | 1-- | 5C+2B | | | | MI | MIRA | |
| KoreaU | 1-- | 5C | GTag | | GDep | Rules | ME | WSJ |

Table 4: Profiles of the participants: GTag=GENIAtagger, MLN=Mark MLS=UMLS SPECIALIST Lexicon/tools, MC=McClosky-Charniak, GDep=Genia Deper rd=Stanford Parser, CBR=Case-Based Reasoning, CM=ConceptMapper.


GENiA
% of systems up to rank using
100
80
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
rank
Full parsing
ML for args
ML for triggers


% of systems up to rank using
100
80
6
4
20
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
rank
Full parsing
ML for args
ML for triggers
Strong support for value of full parsing:
Applied by all systems ranked in top 50%
(Applied by 73% of all systems)


% of systems up to rank using
100
80
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
rank
Full parsing
ML for args
ML for triggers
Support for value of ML in argument detection:
4/5 (80%) systems in top 25%
6/11 (55%) systems in top 50%
13/22 (60%) of all systems




**ML unhelpful for trigger detection (?)**

**1/5 (20%) system in top 25%**

**4/11 (36%) systems in in top 50%**

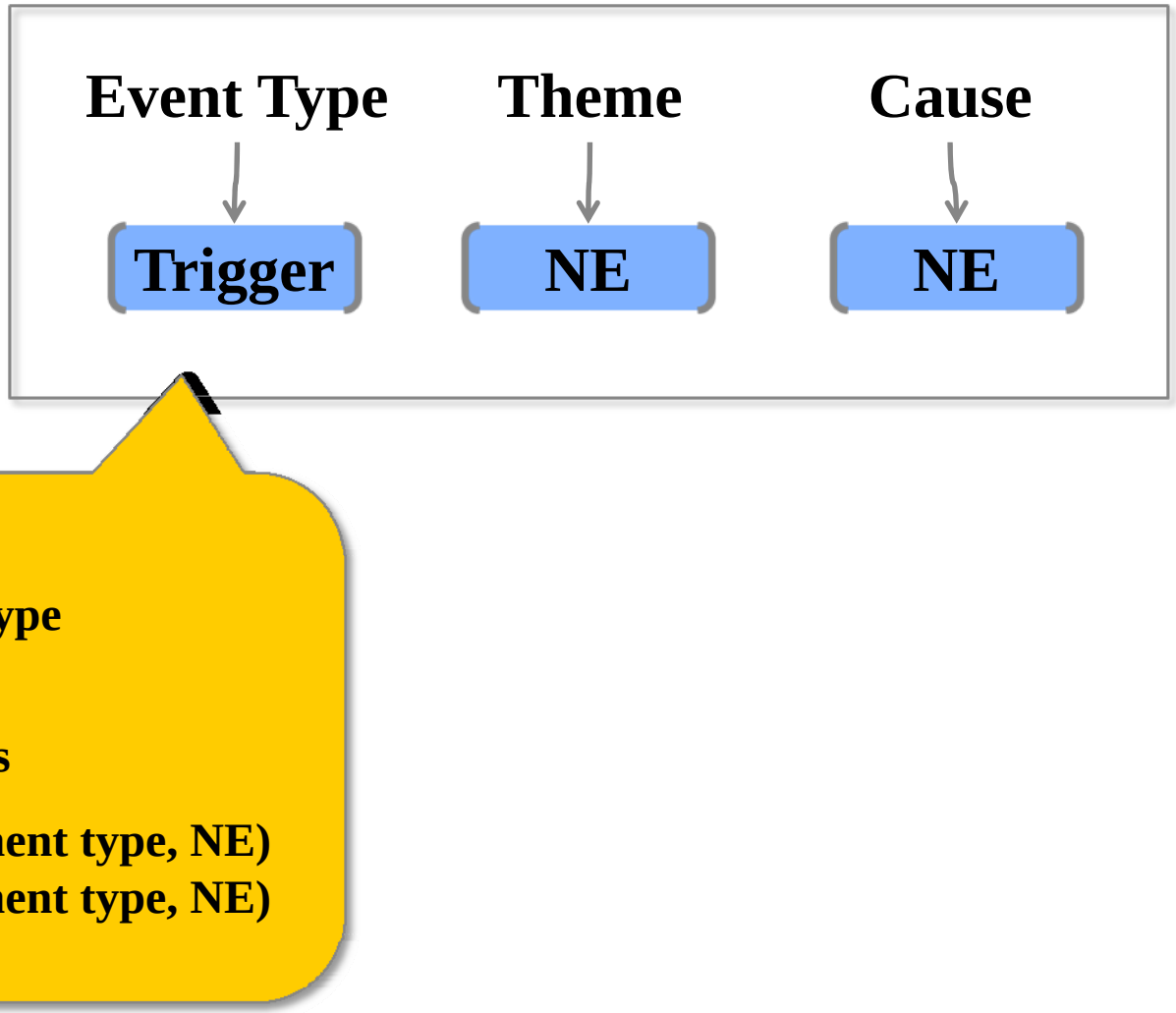
**11/22 (50%) of all systems**

**(difficulty of training with partial annotation?)**

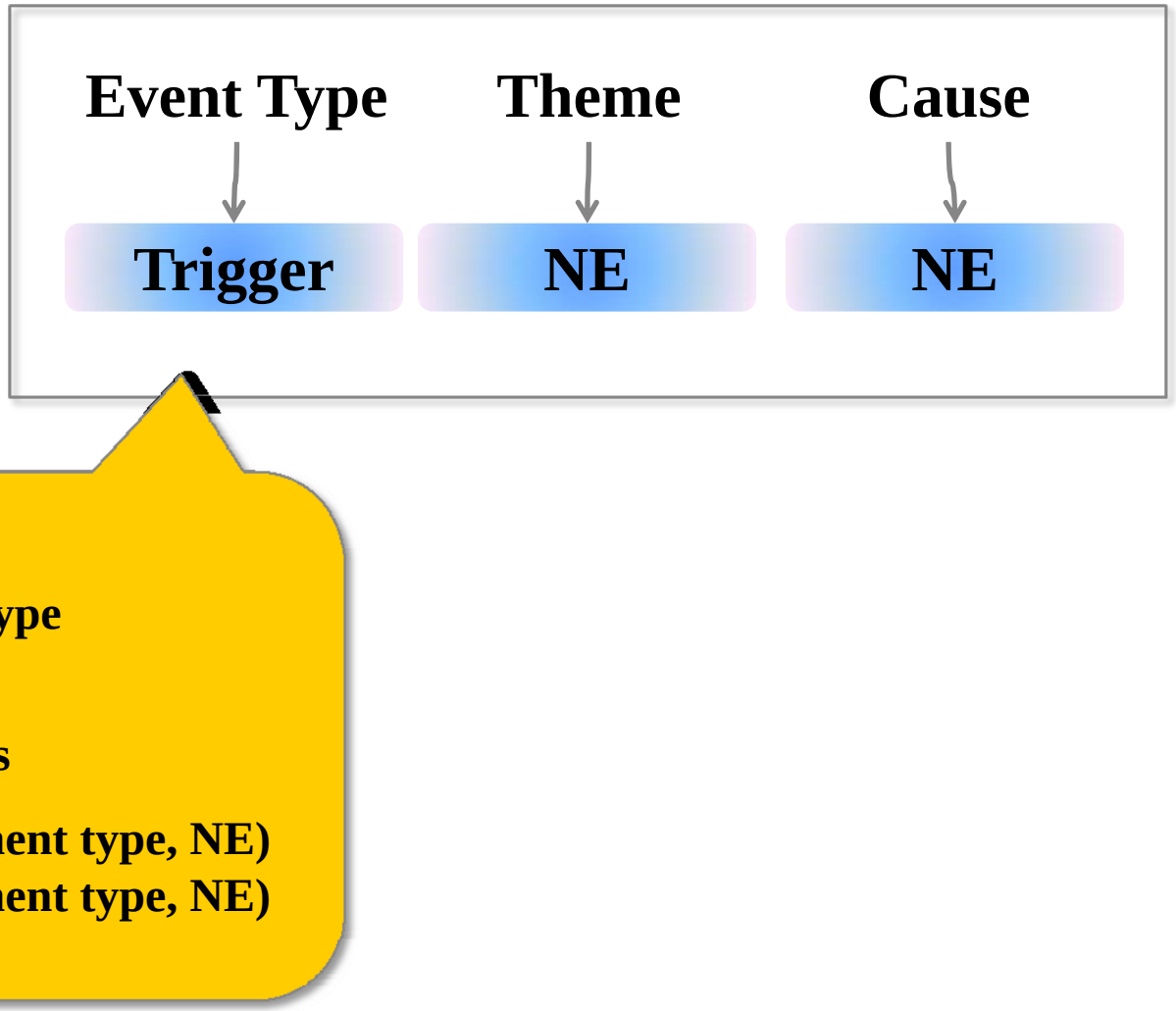
# Evaluation

- Equalities
  - ✓ Event equality
    - ➲ Equality holds between any two events, when
      - (1) The event types are the same,
      - (2) The event triggers are the same, and
      - (3) The arguments are fully matched
  - ✓ Argument equality
    - ➲ Equality holds between any two arguments, when
      - (1) The role types are the same, and
      - (2-1) both are t-entities and equality holds between them, or
      - (2-2) both are events and equality holds between them.
  - ✓ T-entity equality
    - ➲ Equality holds between any two t-entities when
      - (1) the entity types are the same, and
      - (2) the spans are the same.

# Strict Matching

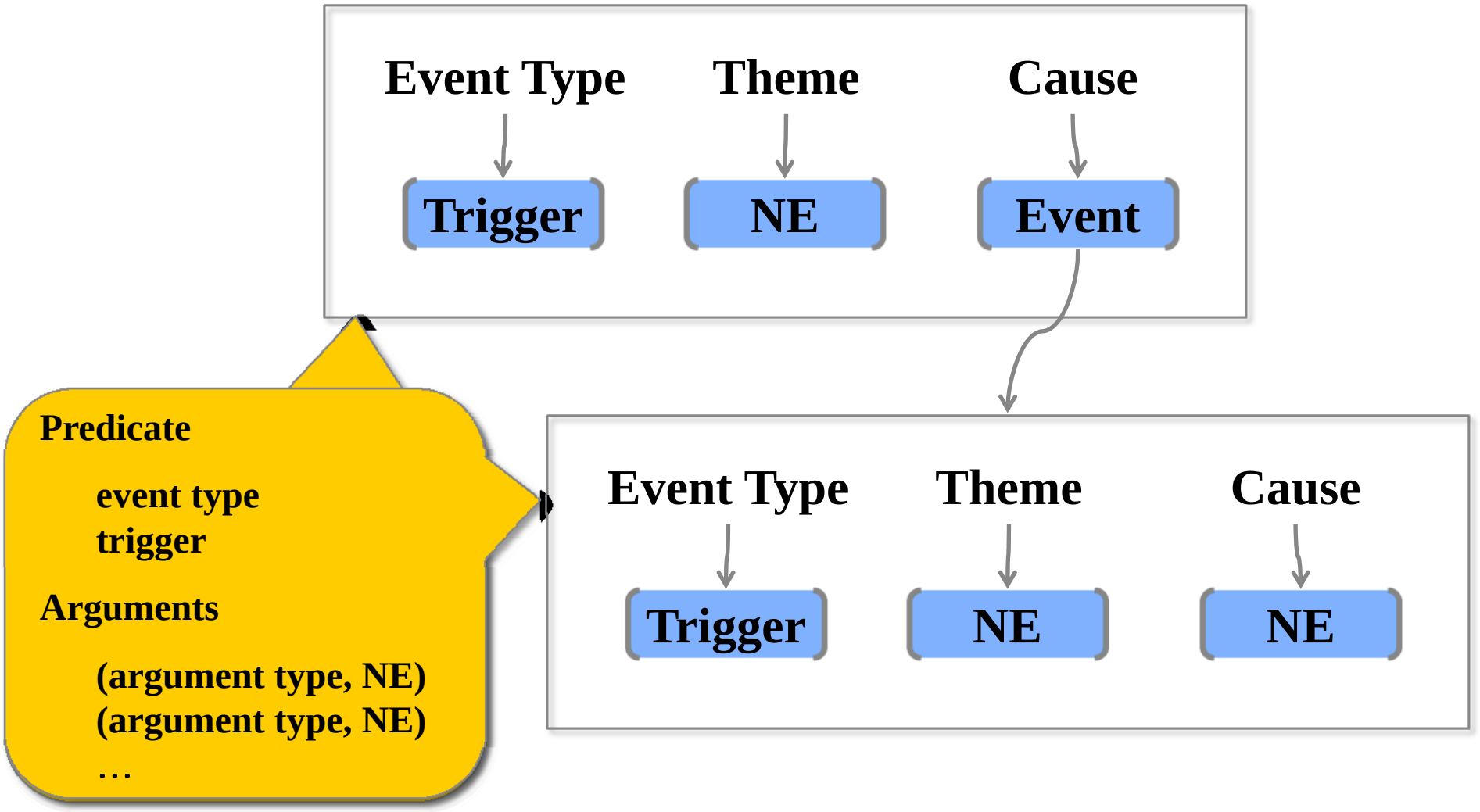


# Approx. Span Matching



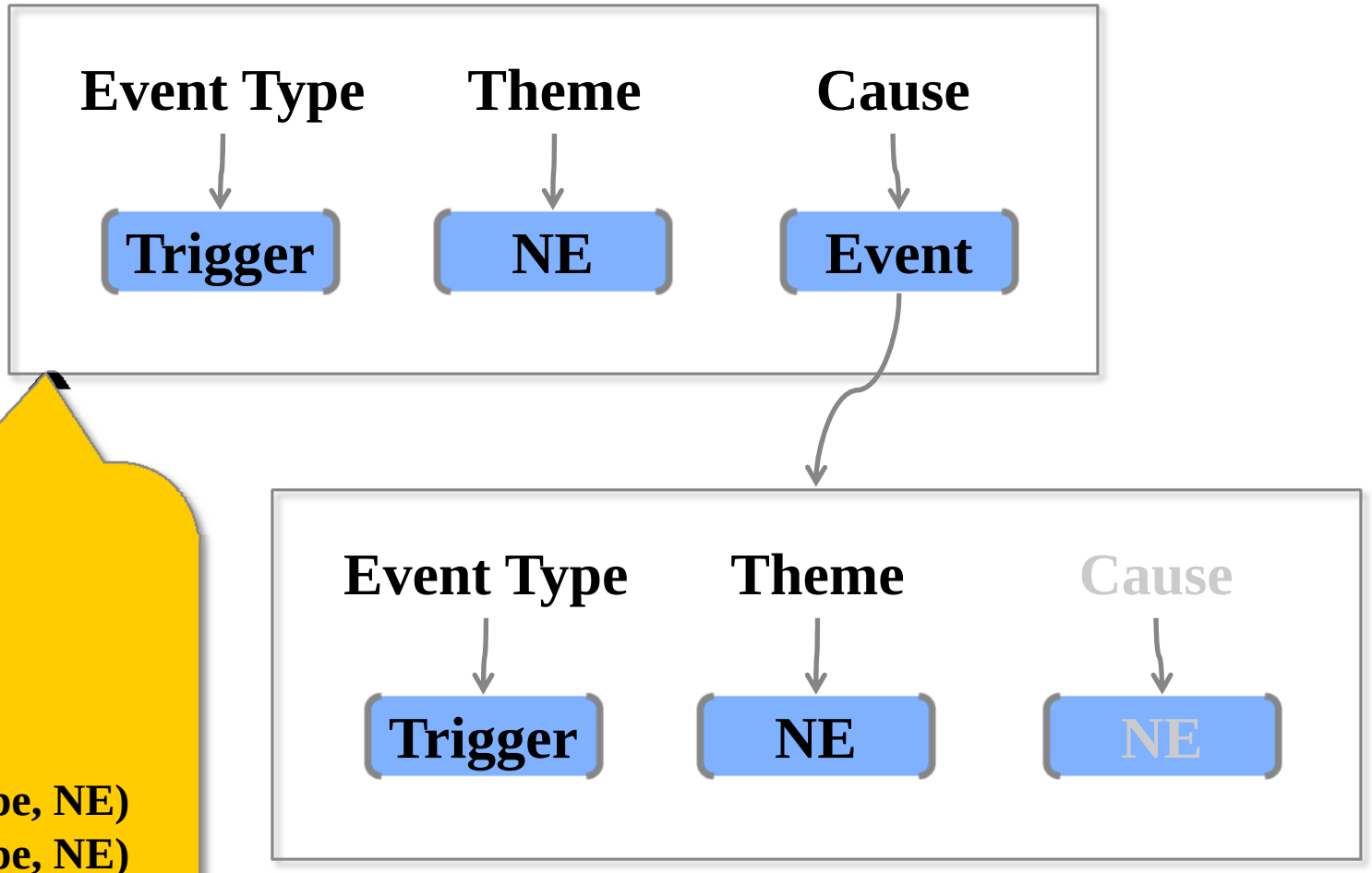
# Strict Matching

## *Regulation type



# Approx. Recursive Matching

## *Regulation type




**Predicate**

event type
trigger

**Arguments**

(argument type, NE)
(argument type, NE)
…

# Official Evaluation

- Evaluation modes
  - ✓ Approximate span matching + Approximate recursive matching

# Results – Task 1

| Team | Simple Event | Binding | Regulation | All |
|---|---|---|---|---|
| UTurku | **64.21 / 77.45 / 70.21** | **40.06 / 49.82 / 44.41** | **35.63 / 45.87 / 40.11** | 46.73 / 58.48 / 51.95 |
| JULIELab | **59.81 / 79.80 / 68.38** | **49.57 / 35.25 / 41.20** | **35.03 / 34.18 / 34.60** | 45.82 / 47.52 / 46.66 |
| ConcordU | **49.75 / 81.44 / 61.76** | 20.46 / 40.57 / 27.20 | **27.47 / 49.89 / 35.43** | 34.98 / 61.59 / 44.62 |
| UT+DBCLS | **55.75 / 72.74 / 63.12** | 23.05 / 48.19 / 31.19 | **26.32 / 41.81 / 32.30** | 36.90 / 55.59 / 44.35 |
| VIBGhent | **54.48 / 79.31 / 64.59** | **38.04 / 38.60 / 38.32** | 17.36 / 31.61 / 22.41 | 33.41 / 51.55 / 40.54 |
| UTokyo | 45.69 / 72.19 / 55.96 | **34.58 / 50.63 / 41.10** | 14.22 / 34.26 / 20.09 | 28.13 / 53.56 / 36.88 |
| UNSW | 45.85 / 69.94 / 55.39 | 23.63 / 37.27 / 28.92 | 16.58 / 28.27 / 20.90 | 28.22 / 45.78 / 34.92 |
| UZurich | 44.92 / 66.62 / 53.66 | 30.84 / 37.28 / 33.75 | 14.82 / 30.21 / 19.89 | 27.75 / 46.60 / 34.78 |
| ASU+HU+BU | 45.09 / 76.80 / 56.82 | 19.88 / 44.52 / 27.49 | 05.20 / 33.46 / 09.01 | 21.62 / 62.21 / 32.09 |
| Cam | 39.17 / 76.40 / 51.79 | 12.68 / 31.88 / 18.14 | 09.98 / 37.76 / 15.79 | 21.12 / 56.90 / 30.80 |
| UAntwerp | 41.29 / 65.68 / 50.70 | 12.97 / 31.03 / 18.29 | 11.07 / 29.85 / 16.15 | 22.50 / 47.70 / 30.58 |
| UNIMAN | 50.00 / 63.21 / 55.83 | 12.68 / 40.37 / 19.30 | 04.05 / 16.75 / 06.53 | 22.06 / 48.61 / 30.35 |
| SCAI | 43.74 / 70.73 / 54.05 | 28.82 / 35.21 / 31.70 | 12.64 / 16.55 / 14.33 | 25.96 / 36.26 / 30.26 |
| UAveiro | 43.57 / 71.63 / 54.18 | 13.54 / 34.06 / 19.38 | 06.29 / 21.05 / 09.69 | 20.93 / 49.30 / 29.38 |
| Team 24 | 41.29 / 64.72 / 50.41 | 22.77 / 35.43 / 27.72 | 09.38 / 19.23 / 12.61 | 22.69 / 40.55 / 29.10 |
| USzeged | 47.63 / 44.44 / 45.98 | 15.27 / 25.73 / 19.17 | 04.17 / 18.21 / 06.79 | 21.53 / 36.99 / 27.21 |
| NICTA | 31.13 / 77.31 / 44.39 | 16.71 / 29.00 / 21.21 | 07.80 / 18.12 / 10.91 | 17.44 / 39.99 / 24.29 |
| CNBMadrid | 50.25 / 46.59 / 48.35 | 33.14 / 20.54 / 25.36 | 12.22 / 07.99 / 09.67 | 28.63 / 20.88 / 24.15 |
| CCP-BTMG | 28.17 / 87.63 / 42.64 | 12.68 / 40.00 / 19.26 | 03.09 / 48.11 / 05.80 | 13.45 / 71.81 / 22.66 |
| CIPS-ASU | 39.68 / 38.60 / 39.13 | 17.29 / 31.58 / 22.35 | 11.86 / 08.15 / 09.66 | 22.78 / 19.03 / 20.74 |
| UMich | 52.71 / 25.89 / 34.73 | 31.70 / 12.61 / 18.05 | 14.22 / 06.56 / 08.98 | 30.42 / 14.11 / 19.28 |
| PIKB | 26.65 / 75.72 / 39.42 | 07.20 / 39.68 / 12.20 | 01.09 / 30.51 / 02.10 | 11.25 / 66.54 / 19.25 |
| Team 09 | 27.16 / 43.61 / 33.47 | 03.17 / 09.82 / 04.79 | 02.42 / 11.90 / 04.02 | 11.69 / 31.42 / 17.04 |
| KoreaU | 20.56 / 66.39 / 31.40 | 12.97 / 50.00 / 20.59 | 00.67 / 37.93 / 01.31 | 09.40 / 61.65 / 16.31 |

# Results – Task 1

| Team | Simple Event | | | |
|---|---|---|---|---|
| UTurku | **64.21 / 77.45 / 70.21** | | | |
| JULIELab | **59.81 / 79.80 / 68.38** | | | |
| ConcordU | **49.75 / 81.44 / 61.76** | 2 | | |
| UT+DBCLS | **55.75 / 72.74 / 63.12** | 23 | | |
| VIBGhent | **54.48 / 79.31 / 64.59** | **38.** | | |
| UTokyo | 45.69 / 72.19 / 55.96 | **34.** | | |
| UNSW | 45.85 / 69.94 / 55.39 | 23. | | |
| UZurich | 44.92 / 66.62 / 53.66 | 30. | | |
| ASU+HU+BU | 45.09 / 76.80 / 56.82 | 19. | | |
| Cam | 39.17 / 76.40 / 51.79 | 12. | | |
| UAntwerp | 41.29 / 65.68 / 50.70 | 12. | | |
| UNIMAN | 50.00 / 63.21 / 55.83 | 12. | | |
| SCAI | 43.74 / 70.73 / 54.05 | 28.82 | | |
| UAveiro | 43.57 / 71.63 / 54.18 | 13.54 / 34.06 / 19.38 | 06.29 / 21.05 / 09.69 | 20.93 / 49.30 / 29.38 |
| Team 24 | 41.29 / 64.72 / 50.41 | 22.77 / 35.43 / 27.72 | 09.38 / 19.23 / 12.61 | 22.69 / 40.55 / 29.10 |
| USzeged | 47.63 / 44.44 / 45.98 | 15.27 / 25.73 / 19.17 | 04.17 / 18.21 / 06.79 | 21.53 / 36.99 / 27.21 |
| NICTA | 31.13 / 77.31 / 44.39 | 16.71 / 29.00 / 21.21 | 07.80 / 18.12 / 10.91 | 17.44 / 39.99 / 24.29 |
| CNBMadrid | 50.25 / 46.59 / 48.35 | 33.14 / 20.54 / 25.36 | 12.22 / 07.99 / 09.67 | 28.63 / 20.88 / 24.15 |
| CCP-BTMG | 28.17 / 87.63 / 42.64 | 12.68 / 40.00 / 19.26 | 03.09 / 48.11 / 05.80 | 13.45 / 71.81 / 22.66 |
| CIPS-ASU | 39.68 / 38.60 / 39.13 | 17.29 / 31.58 / 22.35 | 11.86 / 08.15 / 09.66 | 22.78 / 19.03 / 20.74 |
| UMich | 52.71 / 25.89 / 34.73 | 31.70 / 12.61 / 18.05 | 14.22 / 06.56 / 08.98 | 30.42 / 14.11 / 19.28 |
| PIKB | 26.65 / 75.72 / 39.42 | 07.20 / 39.68 / 12.20 | 01.09 / 30.51 / 02.10 | 11.25 / 66.54 / 19.25 |
| Team 09 | 27.16 / 43.61 / 33.47 | 03.17 / 09.82 / 04.79 | 02.42 / 11.90 / 04.02 | 11.69 / 31.42 / 17.04 |
| KoreaU | 20.56 / 66.39 / 31.40 | 12.97 / 50.00 / 20.59 | 00.67 / 37.93 / 01.31 | 09.40 / 61.65 / 16.31 |

## Simple type events

**Performance range: 30.40~70.21**

**→ Task is not trivial**

**Top systems show acceptable performance**

# Results – Task 1

| Team | Simple Event | Binding | Regulation | All |
|---|---|---|---|---|
| UTurku | **64.21 / 77.45 / 70.21** | **40.06 / 49.82 / 44.41** | **35.63 / 45.87 / 40.11** | |
| JULIELab | **59.81 / 79.80 / 68.38** | **49.57 / 35.25 / 41.20** | **35.03 / 34.18 / 34.60** | |
| ConcordU | **49.75 / 81.44 / 61.76** | 20.46 / 40.57 / 27.20 | **27.47 / 49.89 / 35.43** | |
| UT+DBCLS | **55.75 / 72.74 / 63.12** | 23.05 / 48.19 / 31.19 | **26.32 / 41.81 / 32.30** | |
| VIBGhent | **54.48 / 79.31 / 64.59** | **38.04 / 38.60 / 38.32** | 17.36 / 31.61 / 22.41 | |
| UTokyo | 45.69 / 72.19 / 55.96 | **34.58 / 50.63 / 41.10** | 14.22 / 34.26 / 20.09 | |
| ASU+HU+BU | 45.09 / 76.80 / 56.82 | 19.88 / 44.52 / 27.49 | 05.20 / 33.46 / 09.01 | 21.62 / 62.21 / 32.09 |
| Cam | 39.17 / 76.40 / 51.79 | 12.68 / 31.88 / 18.14 | 09.98 / 37.76 / 15.79 | 21.12 / 56.90 / 30.80 |
| UAntwerp | 41.29 / 65.68 / 50.70 | 12.97 / 31.03 / 18.29 | 11.07 / 29.85 / 16.15 | 22.50 / 47.70 / 30.58 |
| UNIMAN | 50.00 / 63.21 / 55.83 | 12.68 / 40.37 / 19.30 | 04.05 / 16.75 / 06.53 | 22.06 / 48.61 / 30.35 |
| SCAI | 43.74 / 70.73 / 54.05 | 28.82 / 35.21 / 31.70 | 12.64 / 16.55 / 14.33 | 25.96 / 36.26 / 30.26 |
| UAveiro | 43.57 / 71.63 / 54.18 | 13.54 / 34.06 / 19.38 | 06.29 / 21.05 / 09.69 | 20.93 / 49.30 / 29.38 |
| Team 24 | 41.29 / 64.72 / 50.41 | 22.77 / 35.43 / 27.72 | 09.38 / 19.23 / 12.61 | 22.69 / 40.55 / 29.10 |
| USzeged | 47.63 / 44.44 / 45.98 | 15.27 / 25.73 / 19.17 | 04.17 / 18.21 / 06.79 | 21.53 / 36.99 / 27.21 |
| NICTA | 31.13 / 77.31 / 44.39 | 16.71 / 29.00 / 21.21 | 07.80 / 18.12 / 10.91 | 17.44 / 39.99 / 24.29 |
| CNBMadrid | 50.25 / 46.59 / 48.35 | 33.14 / 20.54 / 25.36 | 12.22 / 07.99 / 09.67 | 28.63 / 20.88 / 24.15 |
| CCP-BTMG | 28.17 / 87.63 / 42.64 | 12.68 / 40.00 / 19.26 | 03.09 / 48.11 / 05.80 | 13.45 / 71.81 / 22.66 |
| CIPS-ASU | 39.68 / 38.60 / 39.13 | 17.29 / 31.58 / 22.35 | 11.86 / 08.15 / 09.66 | 22.78 / 19.03 / 20.74 |
| UMich | 52.71 / 25.89 / 34.73 | 31.70 / 12.61 / 18.05 | 14.22 / 06.56 / 08.98 | 30.42 / 14.11 / 19.28 |
| PIKB | 26.65 / 75.72 / 39.42 | 07.20 / 39.68 / 12.20 | 01.09 / 30.51 / 02.10 | 11.25 / 66.54 / 19.25 |
| Team 09 | 27.16 / 43.61 / 33.47 | 03.17 / 09.82 / 04.79 | 02.42 / 11.90 / 04.02 | 11.69 / 31.42 / 17.04 |
| KoreaU | 20.56 / 66.39 / 31.40 | 12.97 / 50.00 / 20.59 | 00.67 / 37.93 / 01.31 | 09.40 / 61.65 / 16.31 |

# Results – Task 2

| Team | Simple Event | Binding | Regulation | All |
|---|---|---|---|---|

**Lesson:**
**Evaluation on one criterion shows something but not everything**

| Team | Site for Phospho.(56) | Site for Binding (21) | AtLoc & ToLoc (65) | (C)Site for Reg. (54) | ALL 2nd Args (196) |
|---|---|---|---|---|---|
| UT+DBCLS | **71.43 / 71.43 / 71.43** | 04.76 / 50.00 / 08.70 | **23.08 / 88.24 / 36.59** | 12.96 / 58.33 / 21.21 | **32.14 / 72.41 / 44.52** |
| UAntwerp | 00.00 / 00.00 / 00.00 | 04.76 / 25.00 / 08.00 | 01.54 /100.00 / 03.03 | **20.37 / 55.00 / 29.73** | 06.63 / 52.00 / 11.76 |
| ASU+HU+BU | 00.00 / 00.00 / 00.00 | 00.00 / 00.00 / 00.00 | 00.00 / 00.00 / 00.00 | 00.00 / 00.00 / 00.00 | 00.00 / 00.00 / 00.00 |
| Team 24 | **55.36 / 93.94 / 69.66** | 04.76 / 20.00 / 07.69 | **21.54 / 66.67 / 32.56** | **24.07 / 72.22 / 36.11** | **30.10 / 76.62 / 43.22** |
| CCP-BTMG | 30.36 /**100.00** / 46.58 | 00.00 / 00.00 / 00.00 | 00.00 / 00.00 / 00.00 | 00.00 / 00.00 / 00.00 | 08.67 /100.00 / 15.96 |
| CNBMadrid | **85.71 / 57.14 / 68.57** | 28.57 / 02.31 / 04.27 | **32.31 / 47.73 / 38.53** | 42.59 / 03.70 / 06.81 | 50.00 / 09.71 / 16.27 |

Detailed evaluation, only considering secondary argument detection

# Results – Task 3

| Team | Negation | Speculation | Mod All | All |
|---|---|---|---|---|
| ConcordU | **14.98 / 50.75 / 23.13** | **16.83 / 50.72 / 25.27** | 15.86 / 50.74 / 24.17 | 32.68 / 60.83 / 42.52 |
| VIBGhent | **10.57 / 45.10 / 17.13** | 08.65 / 15.79 / 11.18 | 09.66 / 24.85 / 13.91 | 30.55 / 49.57 / 37.80 |
| ASU+HU+BU | 03.96 / 27.27 / 06.92 | 06.25 / 28.26 / 10.24 | 05.06 / 27.85 / 08.56 | 19.63 / 59.92 / 29.57 |
| NICTA | 05.29 / 34.48 / 09.17 | 04.81 / 30.30 / 08.30 | 05.06 / 32.26 / 08.74 | 15.95 / 39.66 / 22.75 |
| USzeged | 05.29 / 01.94 / 02.84 | 12.02 / 03.88 / 05.87 | 08.51 / 02.93 / 04.36 | 19.96 / 23.19 / 21.46 |
| CCP-BTMG | 01.76 / 05.26 / 02.64 | 06.73 / 13.33 / 08.95 | 04.14 / 09.94 / 05.84 | 12.33 / 57.40 / 20.30 |

# Ensemble

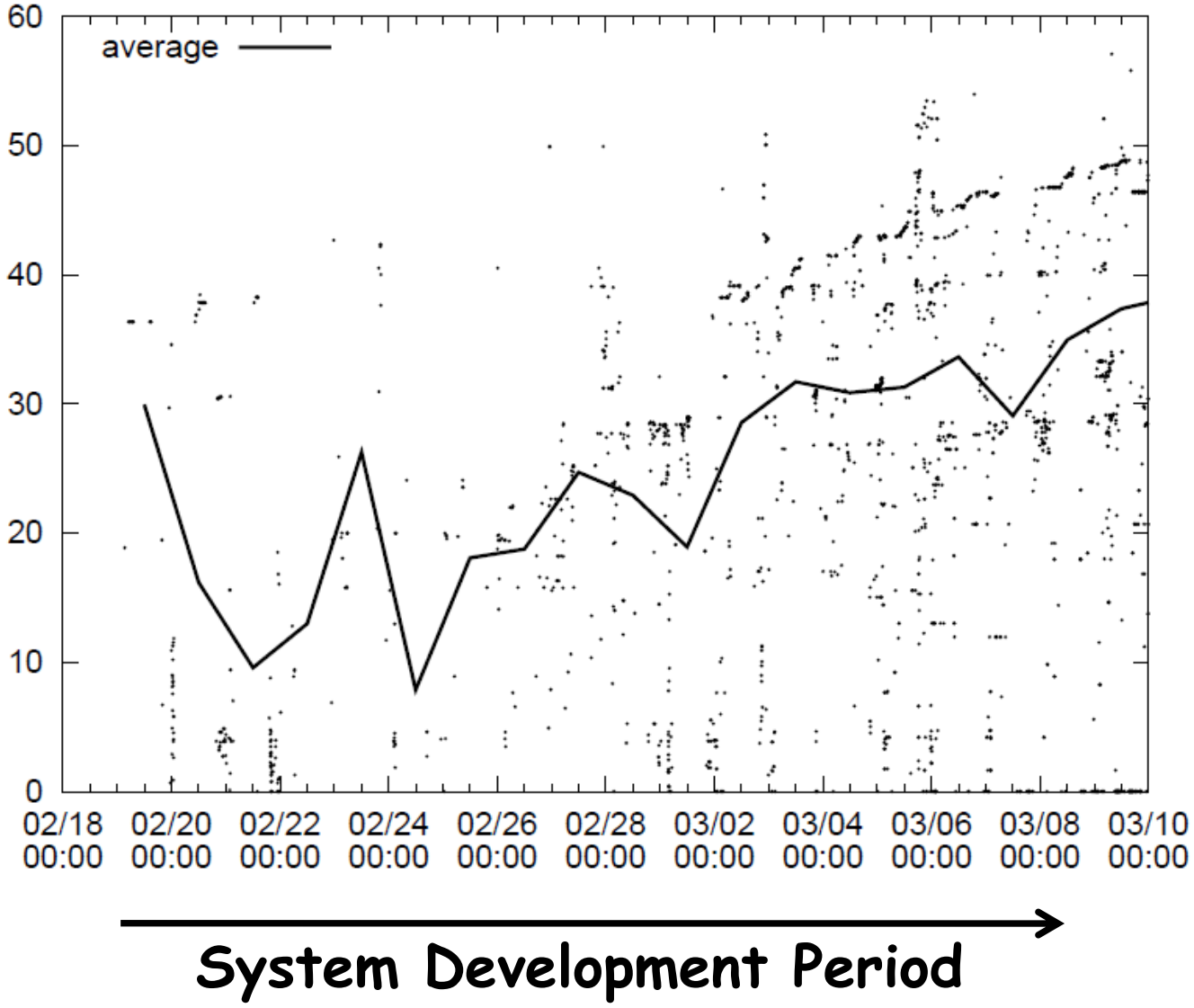
- Experiments for ensemble of top n-systems
  - ✓ By treating the output of each system as a weighted vote
  - ✓ 3 weighting schemes
    - ➲ Equal
    - ➲ Averaged
      - – Weight each vote by the overall f-score of the system.
    - ➲ Event type
      - – Weight each vote by the f-score of the system for the specific event type.
  - ✓ Results

| Ensemble | Equal | Averaged | Event Type |
|---|---|---|---|
| Top 3 | 53.19 | 53.19 | 54.08 |
| Top 4 | 54.34 | 54.34 | 55.21 |
| Top 5 | 54.77 | 55.03 | 55.10 |
| **Top 6** | **55.13** | **55.77** | **55.96** |
| Top 7 | 54.33 | 55.45 | 55.73 |
| Top 10 | 52.79 | 54.63 | 55.18 |

Best result using
- top 6 systems, and
- "Event type" weighting scheme

Improvement over the top result (46.73/58.48/51.95) by 4%.

# Learning Curve



# Conclusions

- BioNLP'09 Shared Task
  - ✓ The first community-wide effort for fine-grained event extraction.
  - ✓ Results
    - ➲ Encouraging
      - 70% f-score for simple events
    - ➲ Reveal remaining challenges
      - Multi-arguments events (Binding)
      - Recursively defined events (Regulation)
      - Causal relation (Regulation)
      - Non-text-bound IE (Negation/speculation)
  - ✓ Successful
    - ➲ active participation
    - ➲ insightful results
  - ✓ Future Works
    - ➲ Further analysis on the results
    - ➲ Collaboration