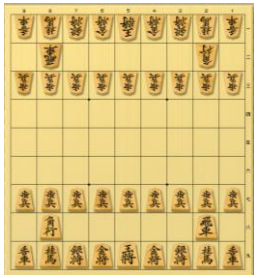
Learning for Information Extraction in Biomedical and General Domains

Makoto MIWA <makoto-miwa@toyota-ti.ac.jp>

Associate Professor, Toyota Technological Institute

December 12, 2016

Self-introduction



Japanese chess “shogi”

I started NLP here

- 2008 Ph.D. thesis on “AI in computer games”
- 2008-2011
  - Researcher, Tsujii laboratory, the University of Tokyo
- 2011-2014
  - Research Associate, NaCTeM, the University of Manchester
  - Visiting researcher, Tsuruoka laboratory, the University of Tokyo
- 2014-
  - Associate Processor, Computational Intelligence laboratory, Toyota Technological Institute (TTI)
  - Visiting Researcher, Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology (AIST)

Main research topics on IE

- PPI extraction (AkaneRE)
- Event extraction (EventMine, PathText)
- Joint entity & relation extraction (JointER, LSTM-ER)

Toyota Technological Institute (TTI)

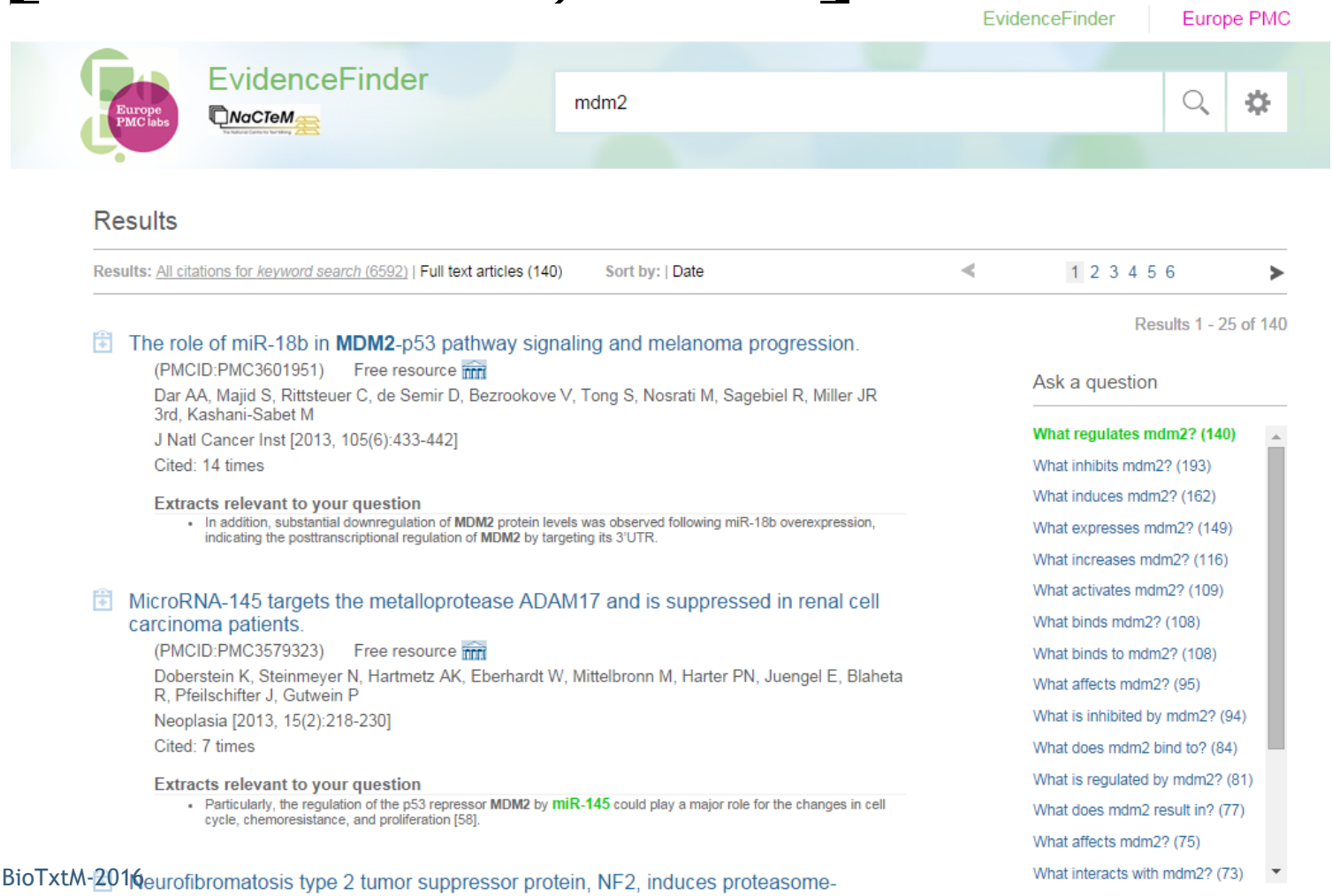
- Located in Nagoya (in between Osaka and Tokyo), Japan
- Founded in 1981 by Toyota Motors Cooperation
- TTI-C (Toyota Technological Institute at Chicago) is a sister university of TTI and TTI-C is founded by TTI in 2003

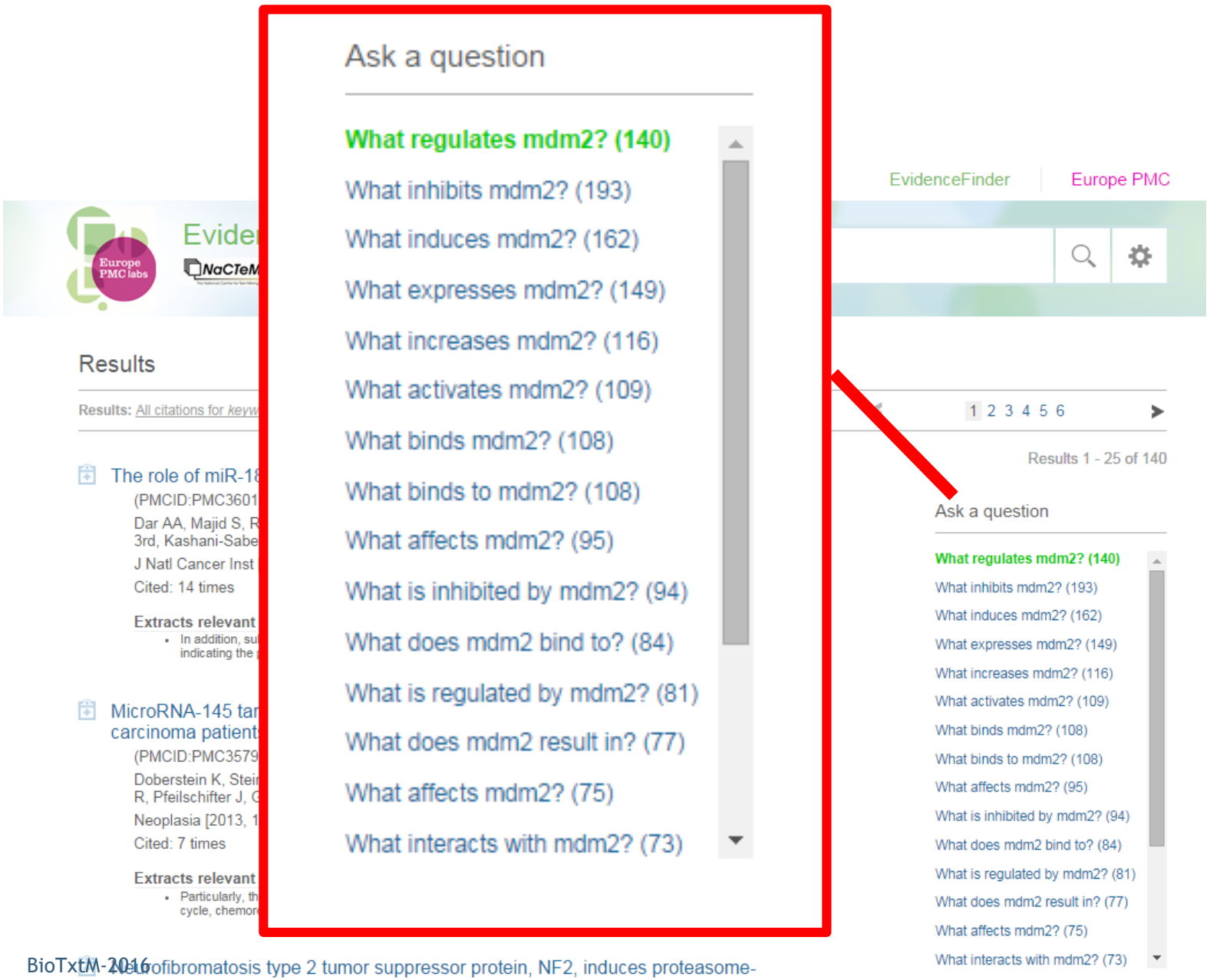
Acknowledgements

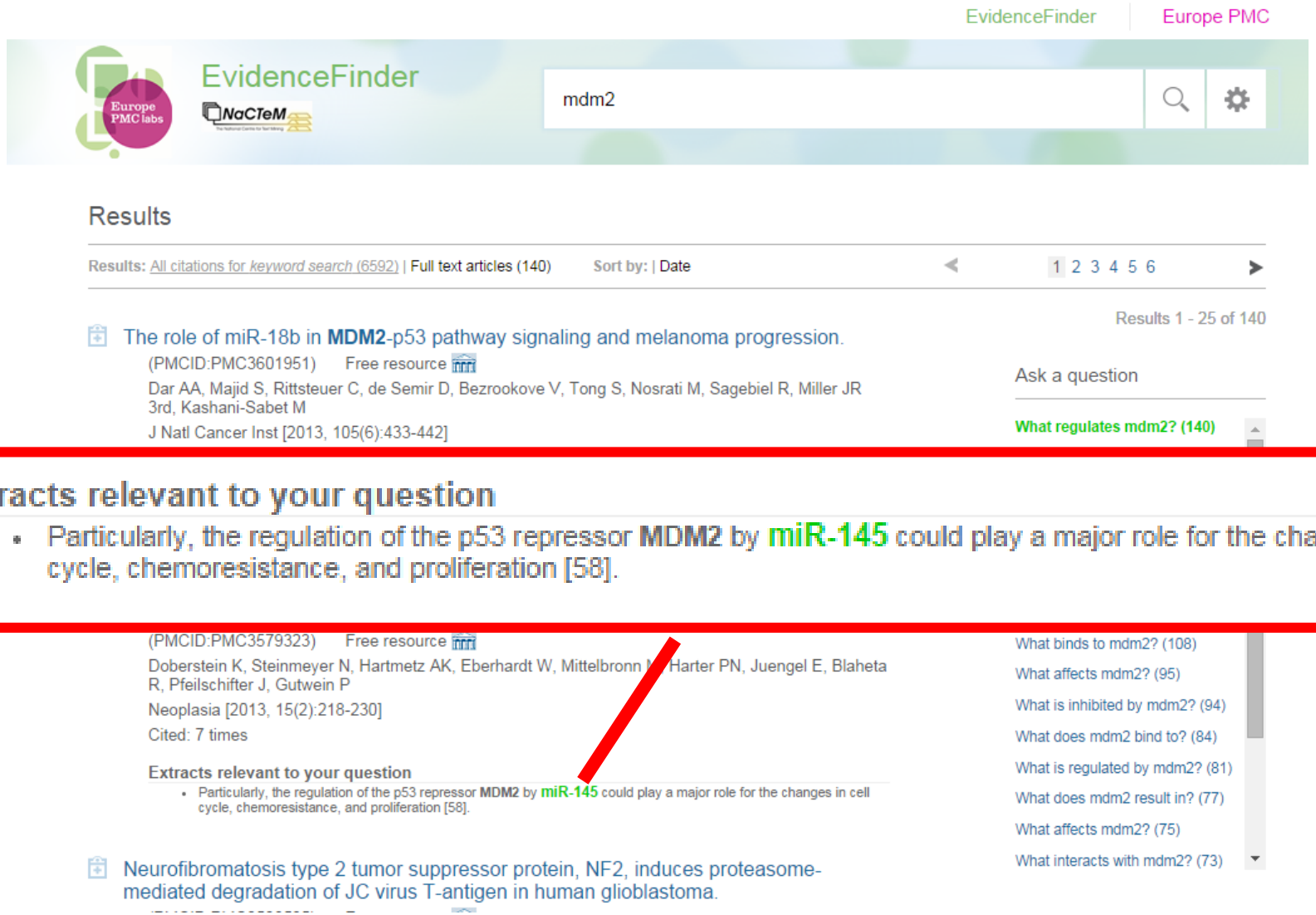
- Most contents are from collaboration with my colleagues at

closed in 2011

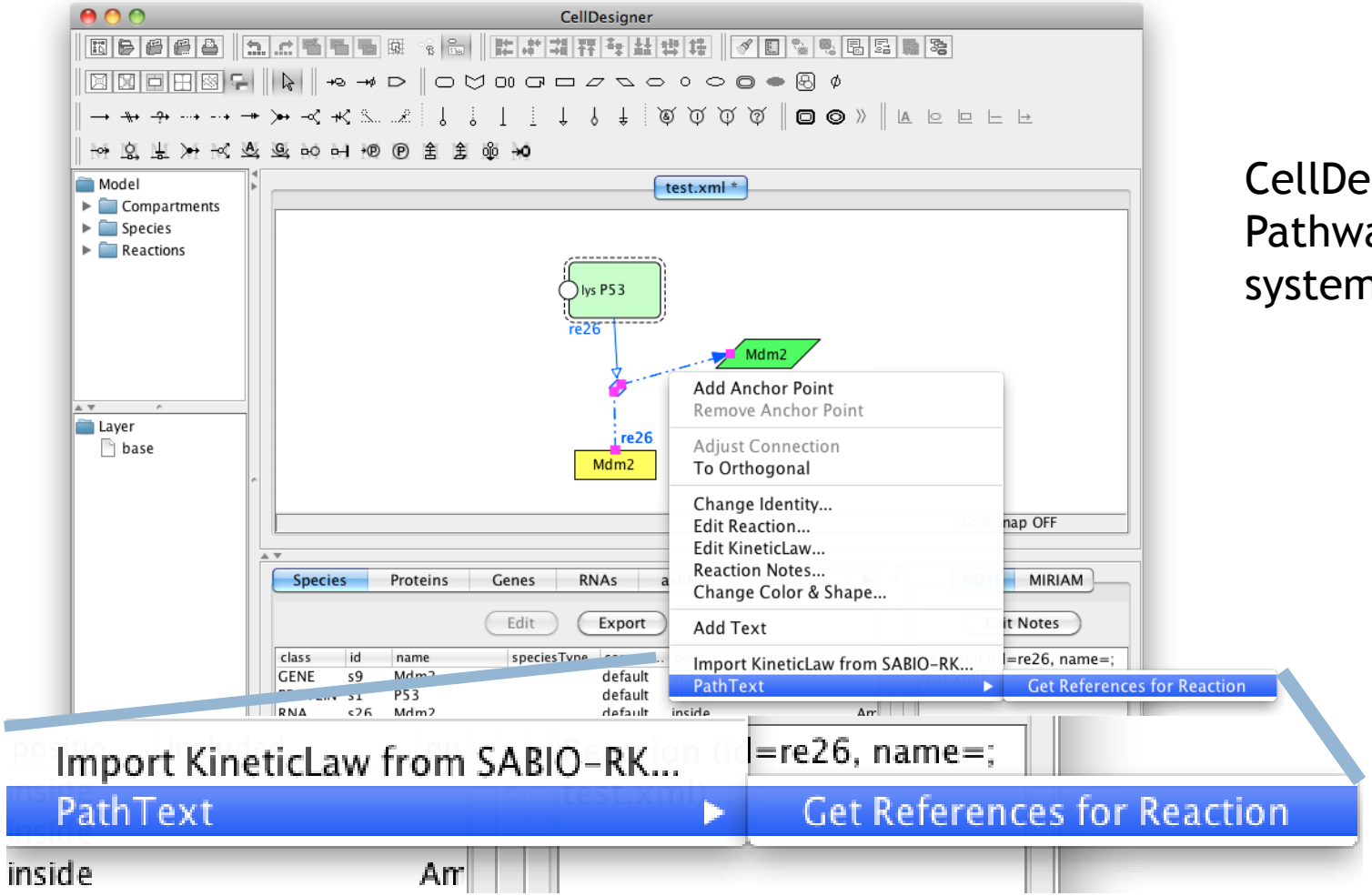
Europe PMC EvidenceFinder [Black et al., 2016]



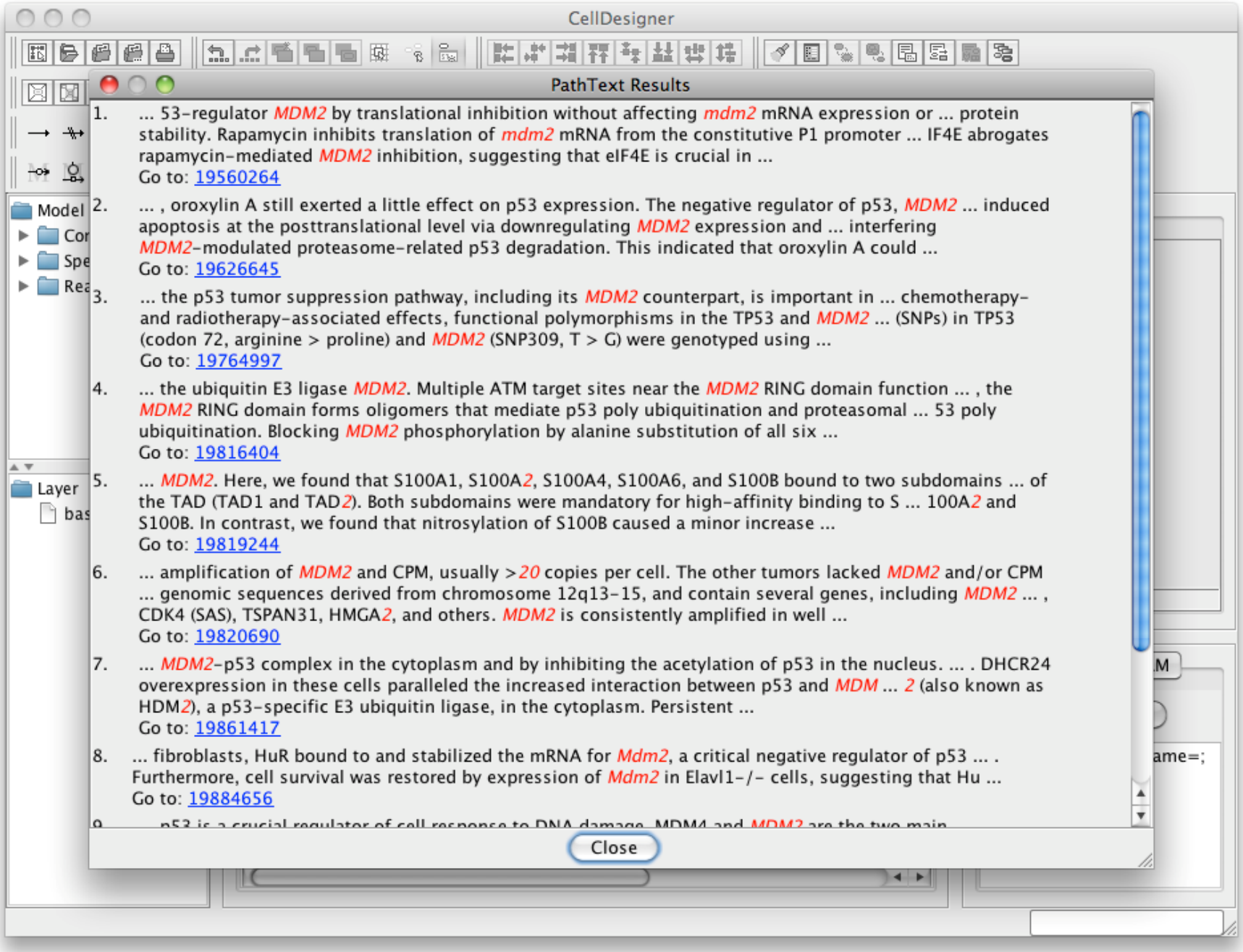

Ask a question
What regulates mdm2? (140)
What inhibits mdm2? (193)
What induces mdm2? (162)
What expresses mdm2? (149)
What increases mdm2? (116)
What activates mdm2? (109)
What binds mdm2? (108)
What binds to mdm2? (108)
What affects mdm2? (95)
What is inhibited by mdm2? (94)
What does mdm2 bind to? (84)
What is regulated by mdm2? (81)
What does mdm2 result in? (77)
What affects mdm2? (75)
What interacts with mdm2? (73)
EvidenceFinder
Europe PMC
Results
1 2 3 4 5 6
Results 1 - 25 of 140
Ask a question
What regulates mdm2? (140)
What inhibits mdm2? (193)
What induces mdm2? (162)
What expresses mdm2? (149)
What increases mdm2? (116)
What activates mdm2? (109)
What binds mdm2? (108)
What binds to mdm2? (108)
What affects mdm2? (95)
What is inhibited by mdm2? (94)
What does mdm2 bind to? (84)
What is regulated by mdm2? (81)
What does mdm2 result in? (77)
What affects mdm2? (75)
What interacts with mdm2? (73)
Neurofibromatosis type 2 tumor suppressor protein, NF2, induces proteasome-mediated degradation of JC virus T-antigen in human glioblastoma.

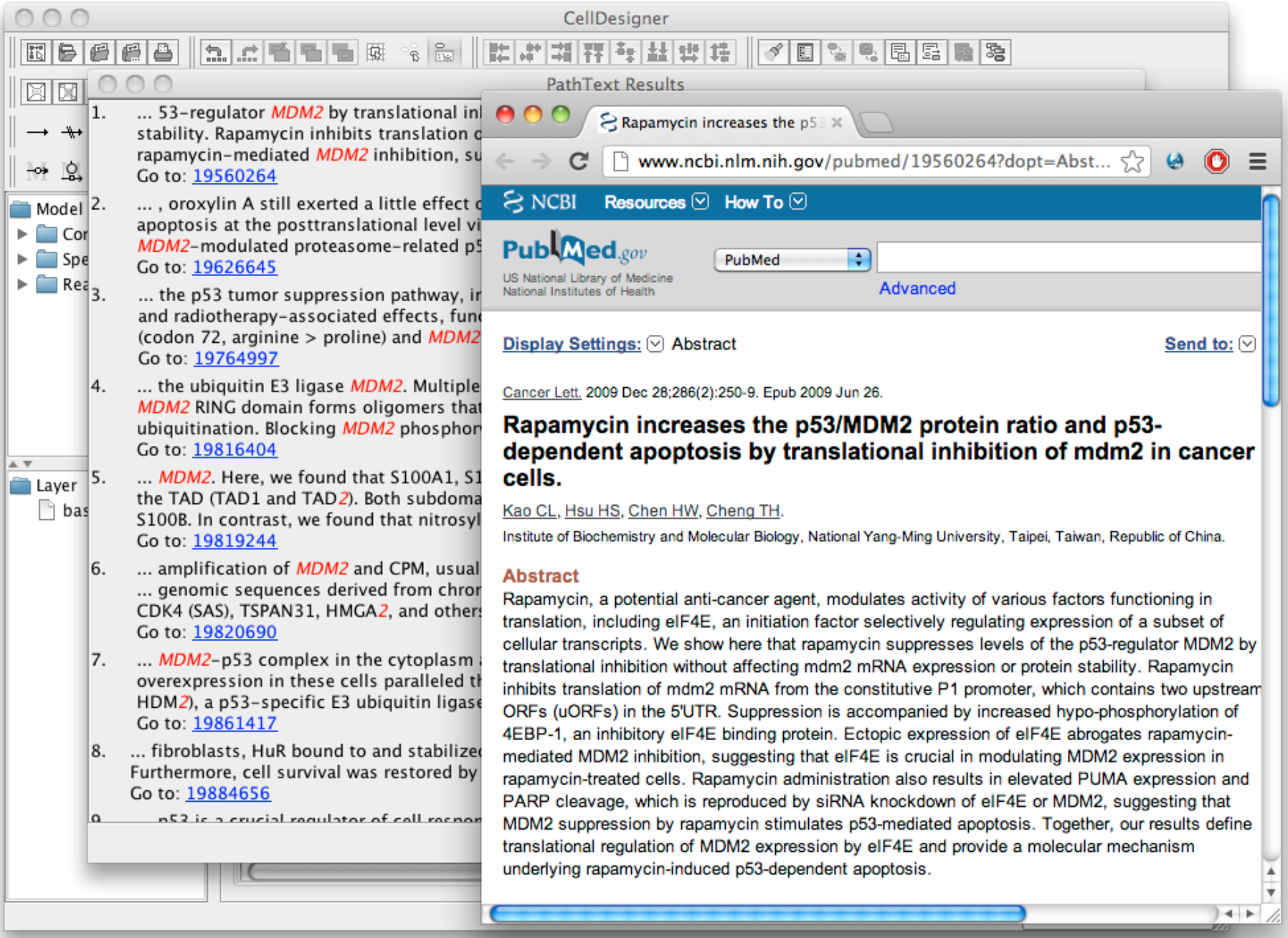

EvidenceFinder
Europe PMC
EvidenceFinder
NaCTeM
mdm2
Results
Results: All citations for keyword search (6592) | Full text articles (140)
Sort by: | Date
1 2 3 4 5 6
Results 1 - 25 of 140
The role of miR-18b in MDM2-p53 pathway signaling and melanoma progression.
(PMCID:PMC3601951) Free resource
Dar AA, Majid S, Rittsteuer C, de Semir D, Bezrookove V, Tong S, Nosrati M, Sagebiel R, Miller JR 3rd, Kashani-Sabet M
J Natl Cancer Inst [2013, 105(6):433-442]
Ask a question
What regulates mdm2? (140)
Extracts relevant to your question
Particularly, the regulation of the p53 repressor MDM2 by miR-145 could play a major role for the changes in cell cycle, chemoresistance, and proliferation [58].
(PMCID:PMC3579323) Free resource
Doberstein K, Steinmeyer N, Hartmetz AK, Eberhardt W, Mittelbronn M, Harter PN, Juengel E, Blaheta R, Pfeilschifter J, Gutwein P
Neoplasia [2013, 15(2):218-230]
Cited: 7 times
Extracts relevant to your question
Particularly, the regulation of the p53 repressor MDM2 by miR-145 could play a major role for the changes in cell cycle, chemoresistance, and proliferation [58].
What binds to mdm2? (108)
What affects mdm2? (95)
What is inhibited by mdm2? (94)
What does mdm2 bind to? (84)
What is regulated by mdm2? (81)
What does mdm2 result in? (77)
What affects mdm2? (75)
What interacts with mdm2? (73)
Neurofibromatosis type 2 tumor suppressor protein, NF2, induces proteasome-mediated degradation of JC virus T-antigen in human glioblastoma.

PathText (CellDesigner Plug-in) [Miwa et al., 2013a]

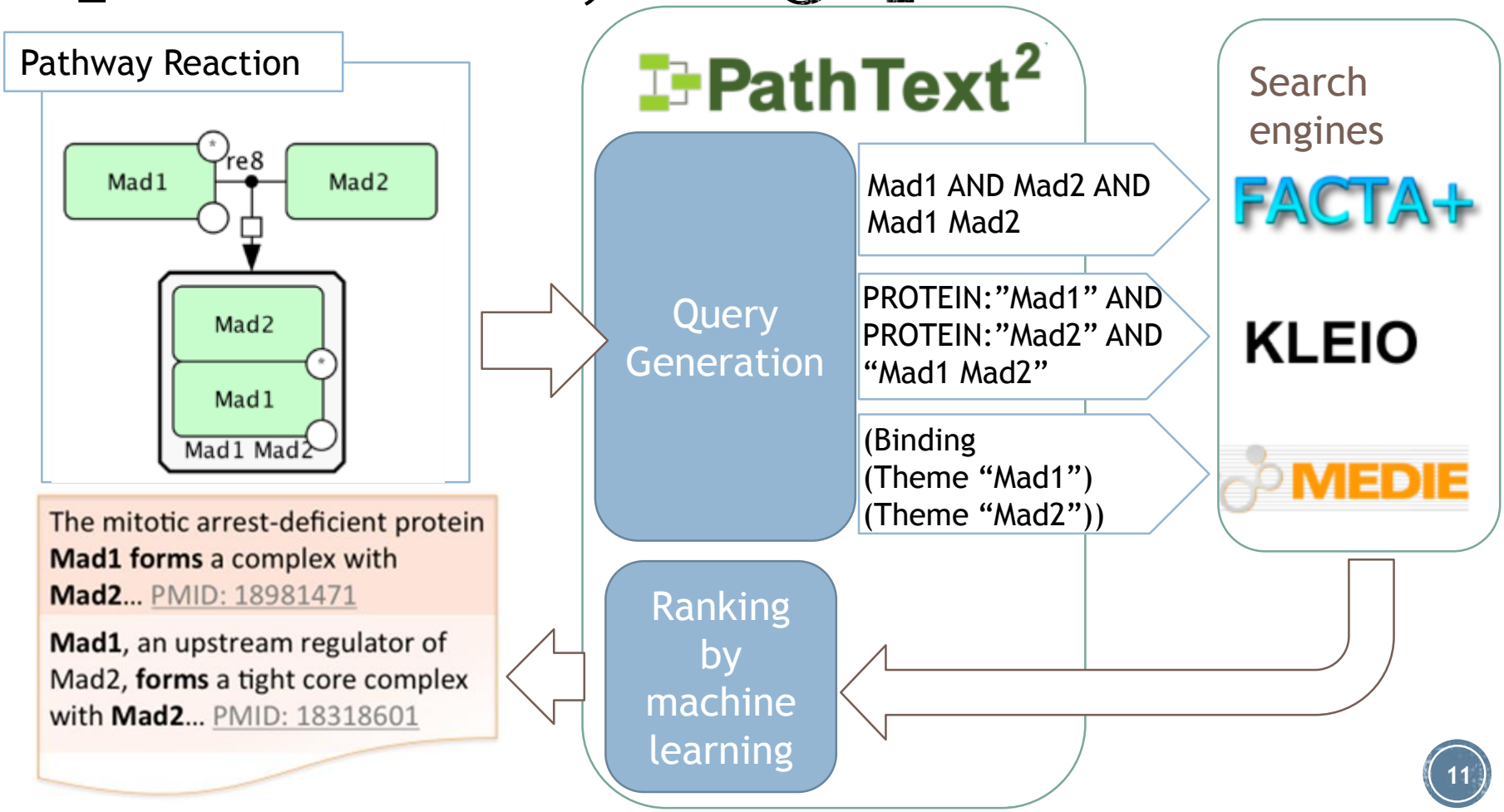


CellDesigner: Pathway editor in systems biology

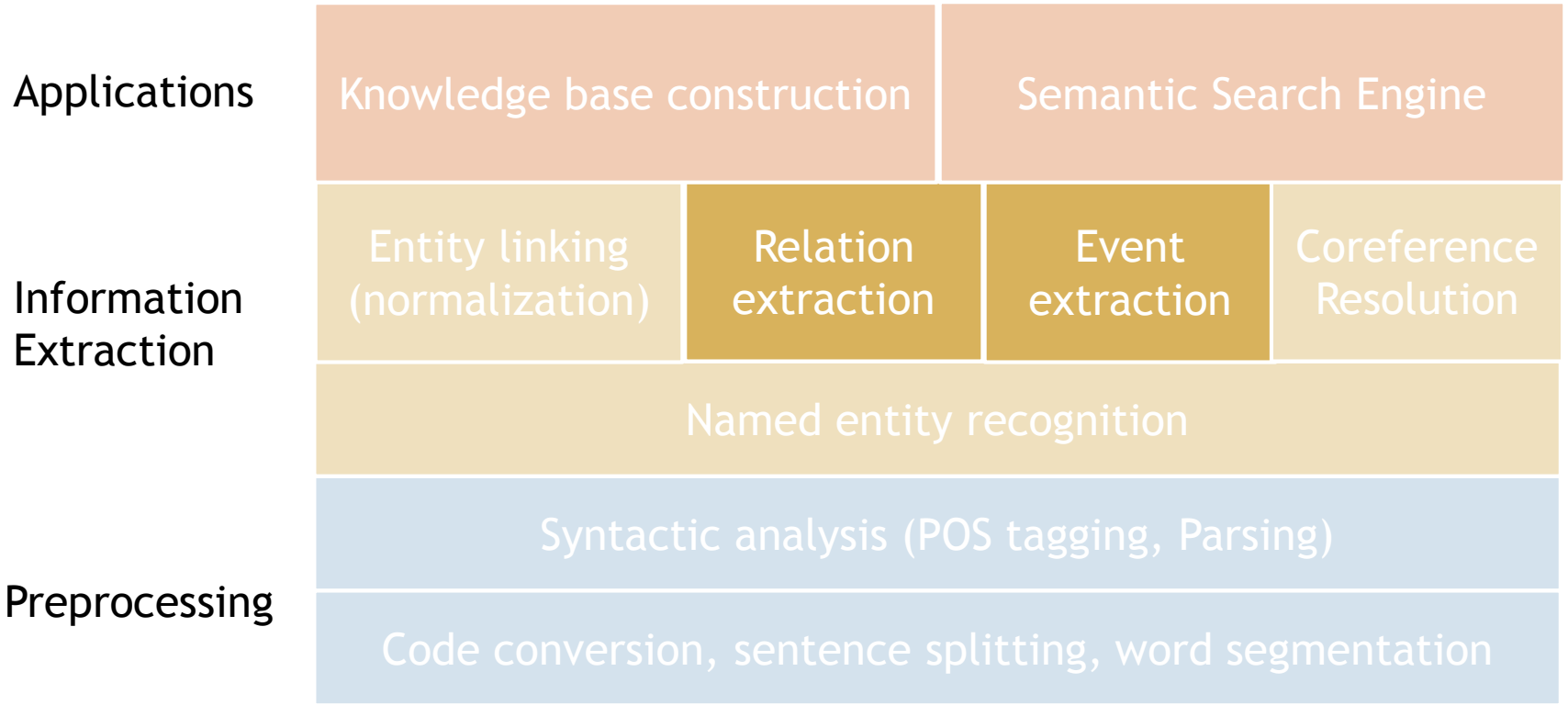

CellDesigner
PathText Results
1. ... 53-regulator MDM2 by translational inhibition without affecting mdm2 mRNA expression or ... protein stability. Rapamycin inhibits translation of mdm2 mRNA from the constitutive P1 promoter ... IF4E abrogates rapamycin-mediated MDM2 inhibition, suggesting that eIF4E is crucial in ...
Go to: 19560264
2. ... , oroxylin A still exerted a little effect on p53 expression. The negative regulator of p53, MDM2 ... induced apoptosis at the posttranslational level via downregulating MDM2 expression and ... interfering MDM2-modulated proteasome-related p53 degradation. This indicated that oroxylin A could ...
Go to: 19626645
3. ... the p53 tumor suppression pathway, including its MDM2 counterpart, is important in ... chemotherapy- and radiotherapy-associated effects, functional polymorphisms in the TP53 and MDM2 ... (SNPs) in TP53 (codon 72, arginine > proline) and MDM2 (SNP309, T > G) were genotyped using ...
Go to: 19764997
4. ... the ubiquitin E3 ligase MDM2. Multiple ATM target sites near the MDM2 RING domain function ... , the MDM2 RING domain forms oligomers that mediate p53 poly ubiquitination and proteasomal ... 53 poly ubiquitination. Blocking MDM2 phosphorylation by alanine substitution of all six ...
Go to: 19816404
5. ... MDM2. Here, we found that S100A1, S100A2, S100A4, S100A6, and S100B bound to two subdomains ... of the TAD (TAD1 and TAD2). Both subdomains were mandatory for high-affinity binding to S ... 100A2 and S100B. In contrast, we found that nitrosylation of S100B caused a minor increase ...
Go to: 19819244
6. ... amplification of MDM2 and CPM, usually >20 copies per cell. The other tumors lacked MDM2 and/or CPM ... genomic sequences derived from chromosome 12q13-15, and contain several genes, including MDM2 ... , CDK4 (SAS), TSPAN31, HMGA2, and others. MDM2 is consistently amplified in well ...
Go to: 19820690
7. ... MDM2-p53 complex in the cytoplasm and by inhibiting the acetylation of p53 in the nucleus. ... . DHCR24 overexpression in these cells paralleled the increased interaction between p53 and MDM ... 2 (also known as HDM2), a p53-specific E3 ubiquitin ligase, in the cytoplasm. Persistent ...
Go to: 19861417
8. ... fibroblasts, HuR bound to and stabilized the mRNA for Mdm2, a critical negative regulator of p53 ... . Furthermore, cell survival was restored by expression of Mdm2 in Elavl1-/- cells, suggesting that Hu ...
Go to: 19884656
Close


CellDesigner
PathText Results
1. ... 53-regulator MDM2 by translational inl
stability. Rapamycin inhibits translation c
rapamycin-mediated MDM2 inhibition, su
Go to: 19560264
2. ... , oroxylin A still exerted a little effect c
apoptosis at the posttranslational level vi
MDM2-modulated proteasome-related p5
Go to: 19626645
3. ... the p53 tumor suppression pathway, ir
and radiotherapy-associated effects, func
(codon 72, arginine > proline) and MDM2
Go to: 19764997
4. ... the ubiquitin E3 ligase MDM2. Multiple
MDM2 RING domain forms oligomers that
ubiquitination. Blocking MDM2 phosphor
Go to: 19816404
5. ... MDM2. Here, we found that S100A1, S1
the TAD (TAD1 and TAD2). Both subdoma
S100B. In contrast, we found that nitrosyl
Go to: 19819244
6. ... amplification of MDM2 and CPM, usual
... genomic sequences derived from chror
CDK4 (SAS), TSPAN31, HMGA2, and others
Go to: 19820690
7. ... MDM2-p53 complex in the cytoplasm
overexpression in these cells paralleled th
HDM2), a p53-specific E3 ubiquitin ligase
Go to: 19861417
8. ... fibroblasts, HuR bound to and stabilized
Furthermore, cell survival was restored by
Go to: 19884656
Model
Cor
Spe
Rea
Layer
bas
Rapamycin increases the p53
www.ncbi.nlm.nih.gov/pubmed/19560264?dopt=Abst...
NCBI Resources How To
PubMed.gov
US National Library of Medicine
National Institutes of Health
PubMed
Advanced
Display Settings: Abstract
Send to:
Cancer Lett. 2009 Dec 28;286(2):250-9. Epub 2009 Jun 26.
Rapamycin increases the p53/MDM2 protein ratio and p53-dependent apoptosis by translational inhibition of mdm2 in cancer cells.
Kao CL, Hsu HS, Chen HW, Cheng TH.
Institute of Biochemistry and Molecular Biology, National Yang-Ming University, Taipei, Taiwan, Republic of China.
Abstract
Rapamycin, a potential anti-cancer agent, modulates activity of various factors functioning in translation, including eIF4E, an initiation factor selectively regulating expression of a subset of cellular transcripts. We show here that rapamycin suppresses levels of the p53-regulator MDM2 by translational inhibition without affecting mdm2 mRNA expression or protein stability. Rapamycin inhibits translation of mdm2 mRNA from the constitutive P1 promoter, which contains two upstream ORFs (uORFs) in the 5'UTR. Suppression is accompanied by increased hypo-phosphorylation of 4EBP-1, an inhibitory eIF4E binding protein. Ectopic expression of eIF4E abrogates rapamycin-mediated MDM2 inhibition, suggesting that eIF4E is crucial in modulating MDM2 expression in rapamycin-treated cells. Rapamycin administration also results in elevated PUMA expression and PARP cleavage, which is reproduced by siRNA knockdown of eIF4E or MDM2, suggesting that MDM2 suppression by rapamycin stimulates p53-mediated apoptosis. Together, our results define translational regulation of MDM2 expression by eIF4E and provide a molecular mechanism underlying rapamycin-induced p53-dependent apoptosis.

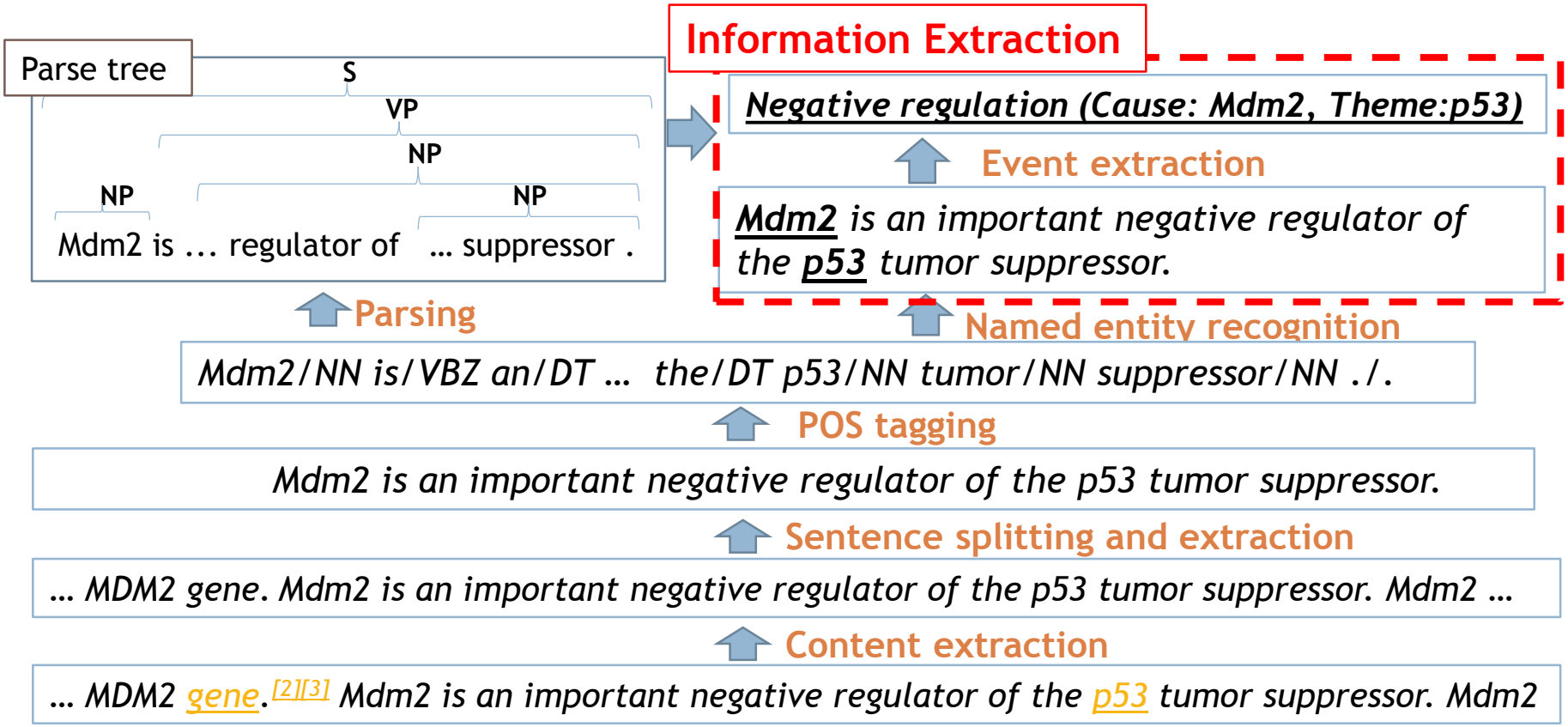
PathText: semantic search engine using extracted events [Miwa et al., 2013a]



Workflow in Information Extraction



Workflow in Information Extraction



http://en.wikipedia.org/wiki/Mdm2

Why you should research information extraction?

- The goal is clear
  - Close to practical applications
- Close to several syntactic and semantic tasks
  - They learn how to use basic NLP tools, e.g., parsers, during development
  - They can move to a variety of related tasks, e.g., NER, co-reference, SRL.
- Complex enough to try a variety of ML methods, but not too complex
  - From simple classifiers to structured learning, and deep learning.
  - Un- or semi-supervised methods can also be considered.
- Clear evaluation metrics and task settings as for recent tasks

➔ Good research targets especially for novices in NLP, TM, and ML

Target problems in this talk

Supervised machine learning (ML) for extracting semantic structures (relations and events) from texts

- Extracting not only terms/words but also ***structures*** *related to them*

⇔Sometime events are single words in knowledge discovery

- Structures need to be linked to some parts of texts as ***evidence***

⇔MUC and knowledge discovery methods often ignore the evidence

- Annotated data and target structures are **manually** built

⇔Distantly (or weakly) supervised methods from knowledge bases may not align with the structures

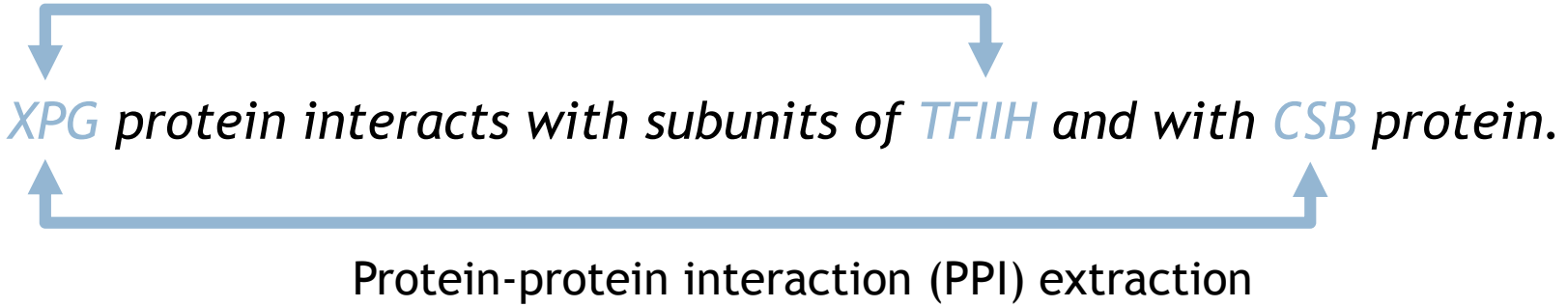
Outline

- Introduction
- Overview of IE tasks
  - General vs Biomedical
- Challenges/considerations in building IE systems
  - Machine learning problems
  - Deep learning
- Our joint entity and relation extraction models
- Conclusion

Overview of IE tasks

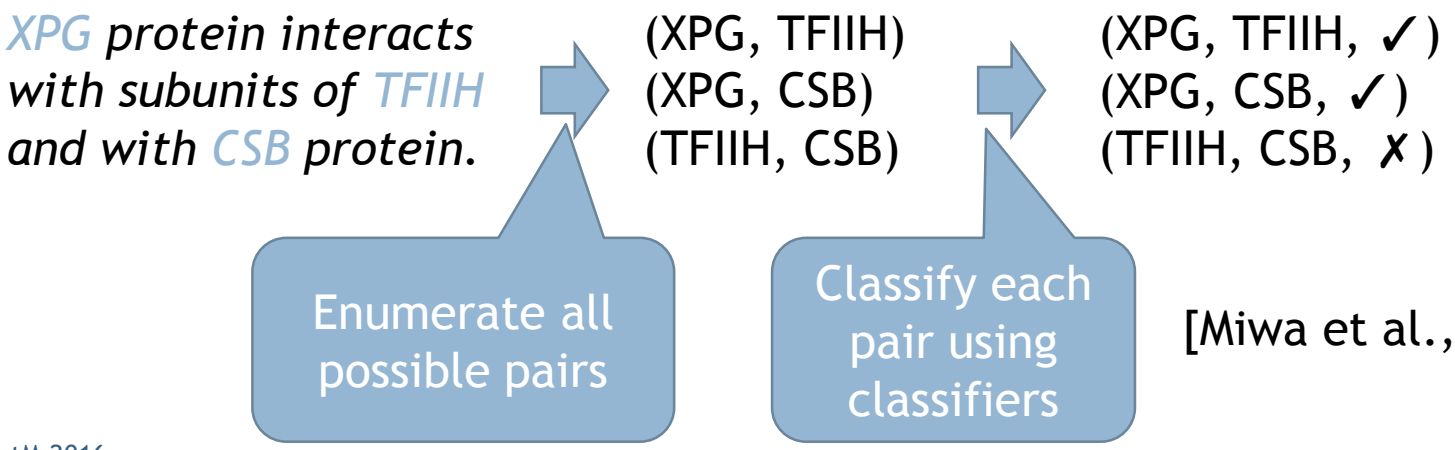
Relation extraction/classification

- Extracting static binary relationships between given entities/nominals
  - Few tasks deal with dynamic and/or n-ary relationships
    - e.g., BB and SeeDev in BioNLP-ST 2016



Classification-based relation extraction (AkaneRE)

- Classifying each target pair as a specific relation or not
  - Pros: easy, fast
  - Cons: no consideration on interactions between relations
- Pairs are represented as *feature vectors* or with multiple *kernels*.



[Miwa et al., 2009]

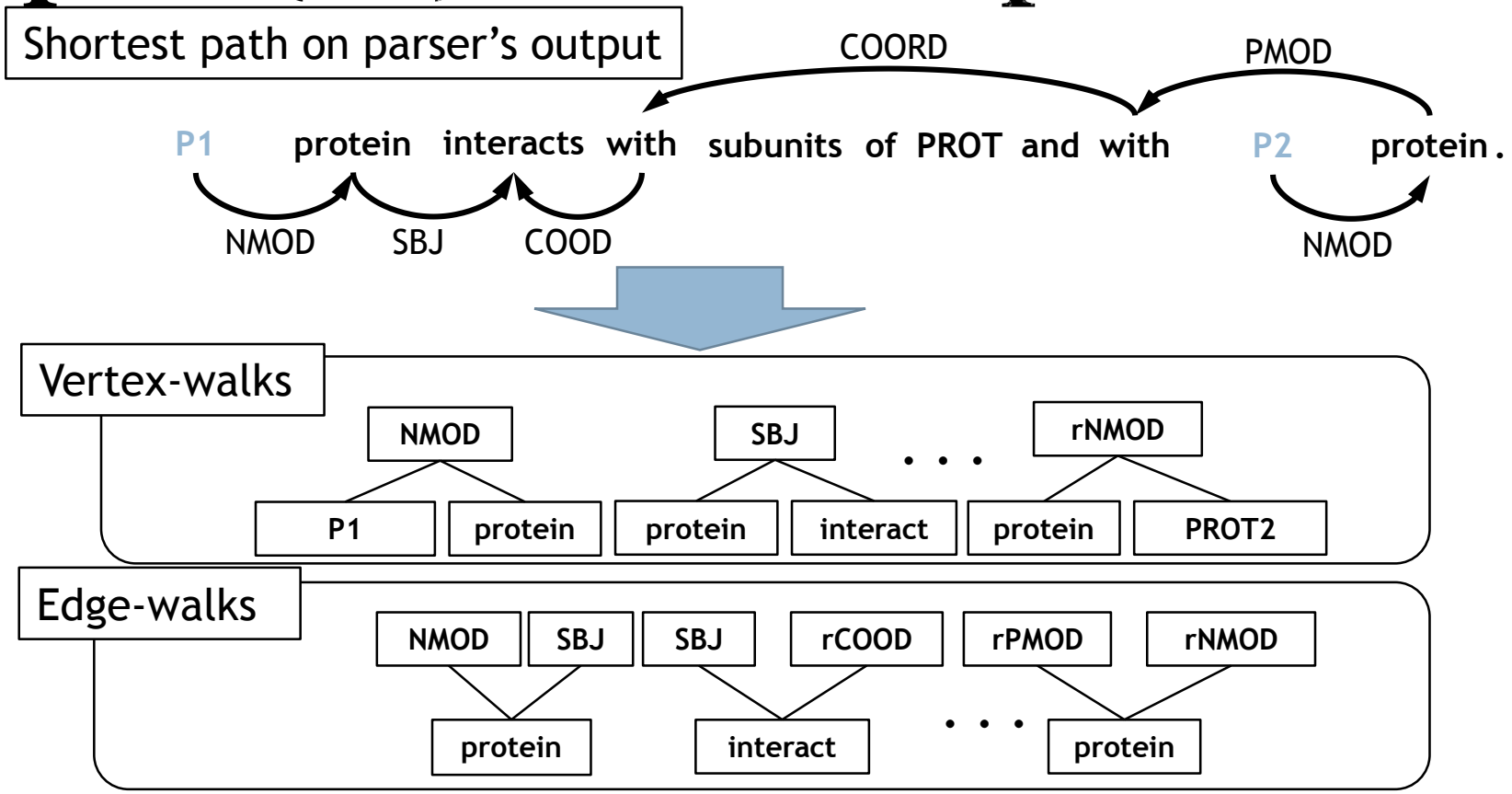
Example of features for relation extraction: bag-of-words (BoW)

P1 protein interacts with multiple subunits of PROT and with P2 protein .

position	features (lemma, frequency, position to a pair)
Before the pair	-
in the Middle of the pair	PROT_M:1, and_M:1, interact_M:1, multiple_M:1, of_M:1, protein_M:1, subunit_M:1, with_M:2
After the pair	protein_A:1

- Features are usually designed as pair-dependent
- Bio-tasks often blind named entities

Example of features for relation extraction: shortest path (SP) between a pair



Other common features/kernels

- Lexical features
  - Character n-grams, part-of-speech
- Contextual features
- Kernels
  - Shortest path kernel
  - Tree kernels
  - Graph kernel

[Miwa et al., 2009]

System	F1-scores on AIMed (%)
AkaneRE (BoW+SP)	63.2
AkaneRE (BoW+SP+Graph)	64.2

- BoW+SP produces close to the state-of-the-art performance (if we ignore deep learning-based methods)

Major relation extraction tasks

Biomedical | General

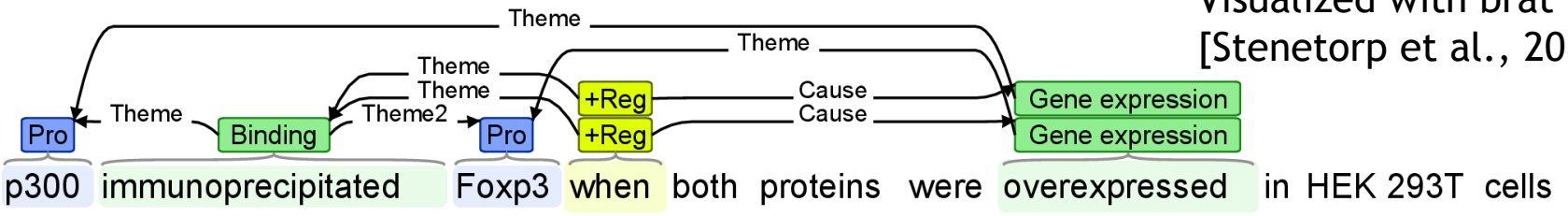
	AIMed (PPI)	DDI-Extraction 2013	ACE2005	SemEval-2010 Task 8
#types	1	4	6 (18 subtypes)	9
directed?	No	No	Yes	Yes
participants	NEs (1 type)	NEs (4 types)	Entities (7 types)	Nominals
Eval. target	relations	typed relations	relations, types	types
metric	CV macro-F1, AUC	Type macro-F1	micro-F1	Type macro-F1

- In Bio domain, relations are often undirected and defined between NEs
  - NEs are often *blinded* in biomedical relation extraction
- SemEval-2010 Task 8 is pure relation classification task (one pair/sentence)
- The differences are minor and same classification methods are applicable, but evaluation metrics are different among tasks
  - Note: evaluation metrics for PPI and ACE2005 are not always same among papers

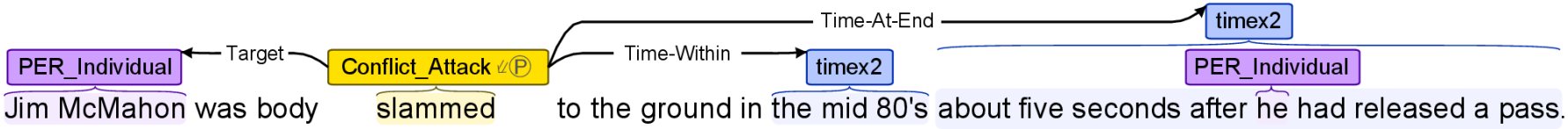
Event extraction

- Extracting dynamic (given) entity state changes and their relationships
  - An event usually consists of a trigger and their typed arguments

Visualized with brat
[Stenetorp et al., 2012]



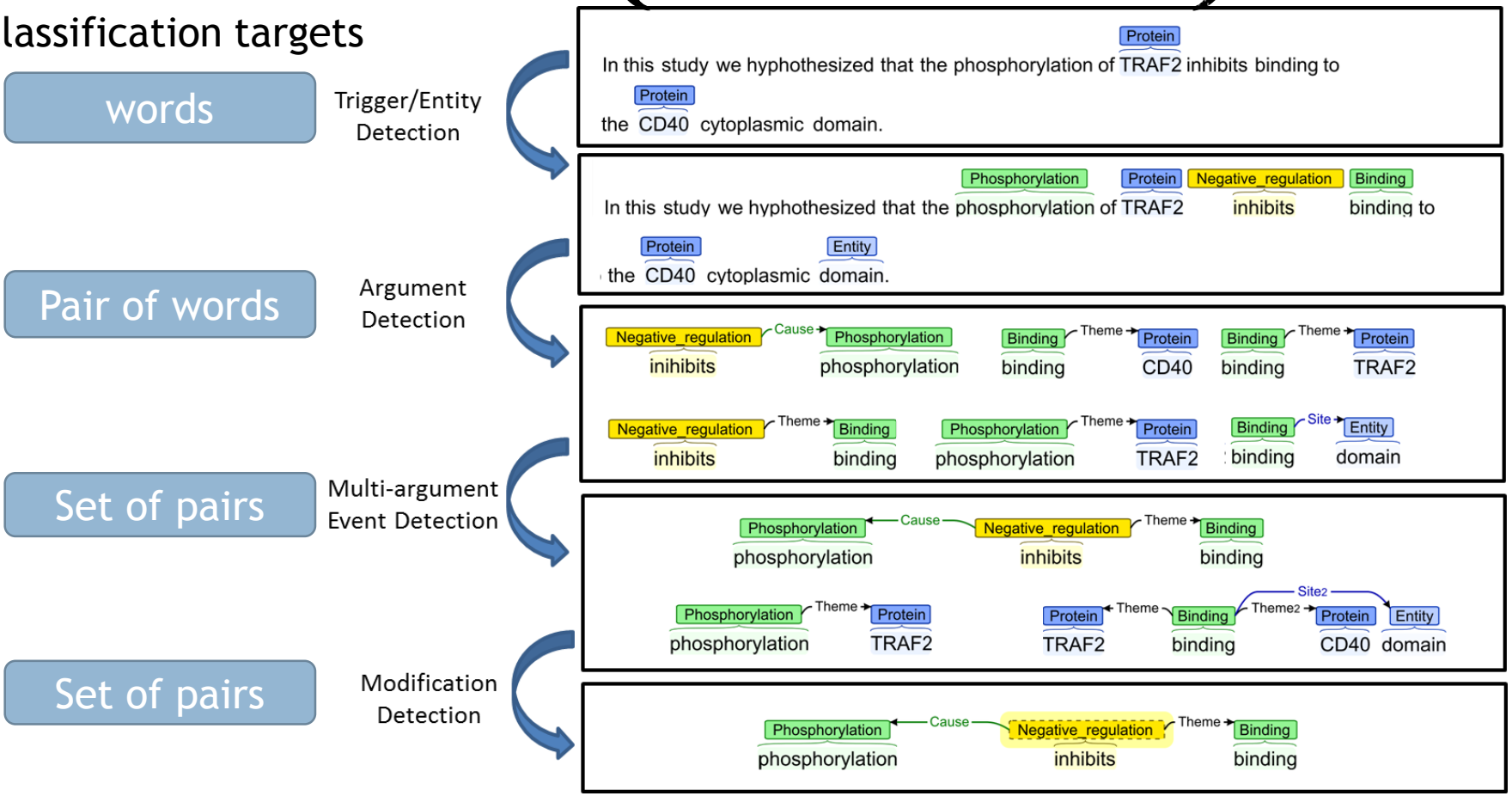
BioNLP Shared Task 2013 GENIA (ID: PMC-1447668-08-Results)



ACE 2005 (ID: MARKBACKER 20041220.0919)

Pipeline-based event extraction (EventMine)

Classification targets



[Miwa et al., 2012]

Comparison between ACE2005 (General) and GENIA2013 (Bio) [Miwa et al., 2014a]

		ACE2005	GENIA2013
type	#event types	8 types / 33 subtypes	13 types
	#entity types	13 types / 53 subtypes	2 types
	#argument role types	35 types	7 types
	argument types	Entity/Nominal/Value/Time	Entity
structure	Max # of arguments	11	4
	Nested events	None	Possible
	Overlaps of events	None	Possible
	Correspondences of arguments	None	Possible

- ACE defines more entity types with a hierarchy

⇔ GENIA in shared task has fewer, flat types

- ACE events have more arguments with flat structure

Comparison between ACE2005 (General) and GENIA2013 (Bio)

		ACE2005	GENIA2013
task	Entity	Given	Partially given
	Entity name	Given	Not available
	Entity attributes	Given	Not available
	Entity coreference	Given	Not given
	Evaluation	Trigger/Role	Event (nested events are broken down)

- ACE provides rich entity information and performs break-down evaluation (due to many arguments)
- Since definitions on event structures are different (flat vs nested), flat ACE event extraction methods are not directly applicable to nested GENIA 2013 events
- Both employs micro-averaged F1 scores

Application of biomedical event extraction system EventMine to general events

F1-scores (%)

	GENIA2013 (Bio)	ACE05 (General)
EventMine	52.71	52.1
Li et al. (2013) (General)	-	52.7
EVEX (Bio)	50.97	-

- EventMine was originally developed for BioNLP tasks, but it also performs well on ACE05
  - Little domain knowledge
  - Not over-tuned to some specific domain
- Biomedical event extraction system can be easily applied to general domain tasks

Summaries in overview

- Basic relation/event extraction systems can be built using classifiers
- Building general IE systems on both bio- and general tasks is not so difficult if we consider the differences including
  - blinded named entities
  - directed relations, and
  - nested events
- Discussions
  - Why do we need to reinvent the methods for almost the same tasks?
  - Can we share strong baseline systems?

Challenges / considerations in building IE systems

Challenges in relation / event extractions (1)

- Ambiguity in analyzing phrase structures
  - >80% of core arguments in GENIA events are covered by the subjects or direct/indirect objects of trigger verbs [Nhung et al., 2015]
  - Disambiguation needs knowledge on contexts outside of the target sentences, and some of them relate to coreference
    - Coreference seems to be a major source of low performance, but the performance on ACE05 is not high with coreference information

the induction of IL-10. ➔ Gene expression
the induction of IL-10 production by Th1 cells ➔ Regulation

Challenges in relation / event extractions (2)

- Variety in expressions

IL-10 production
The induction of IL-10
ability to express IL-10
IL-10-producing Th1 cells

tyrosine phosphorylation of STAT1
phosphorylation of STAT1 on tyrosine
STAT1 tyrosine phosphorylation
phosphorylation of tyrosine on STAT1
phosphorylation on tyrosine of STAT1

- ML-based systems entrust these to features and ML methods
- Other enhancements are also important for performance

➔ I will overview problems and their practical solutions for high performance system, to get to the research baseline

Machine learning-related problems

- Highly imbalanced problems
  - Most pairs are not related (negative)
- Different types of features with different scales
  - E.g., binary features, count-based features, length-based features
- Many features but relatively few instances (compared to other simple tasks)
  - NEs (2,000 abstracts in NLPBA2004) or parsing (~2,500 stories in PTB)
  - Relation/event corpora are less consistent than NE corpora *in nature*
- Pipeline-specific problems

Alleviating imbalanced problems

- High accuracy (ML objectives) does not always mean high F-scores (metrics)

	#positive instances	#negative instances
AIMed	1,000	4,832
DDI 2013 (train)	4,020	23,772

- Rule-based filters of negative instances [Chowdhury et al., 2012]
  - Easy and simple rules work well on some tasks
  - Rules depend on the task, and may not be generalized well
- Weighting / Under-sampling frequent-type instances [Miwa et al, 2012]
  - Preparing two hyper-parameters for positive and negative instances

$$\min_w \frac{1}{2} R(w) + C_p \sum_p L(w, x_p) + C_n \sum_n L(w, x_n)\,, L\text{: loss}, R\text{: regularizer}$$

Robustness against different scales of values

- Larger values are strongly regularized, and they are not learned well

$$\min_w \frac{1}{2}\|\boldsymbol{w}\| + \sum_i L(\mathbf{w}, x_i)$$

- Normalization [Miwa et al., 2009]
  - Normalizing each sub-vector (n-gram, dependency, etc.) of a feature vector, and then globally normalizing the entire vector
    - $\boldsymbol{x} = \frac{x'}{\|x'\|}, \boldsymbol{x}' = \left(\frac{x_1}{||x_1||}, \frac{x_2}{||x_2||}, \ldots.\right)$
  - c.f. “unit kernel” in kernel-based methods, “batch normalization” and “layer normalization” in deep learning
- Scaling
  - Scaling each value with its maximum absolute value

Addressing many features problems

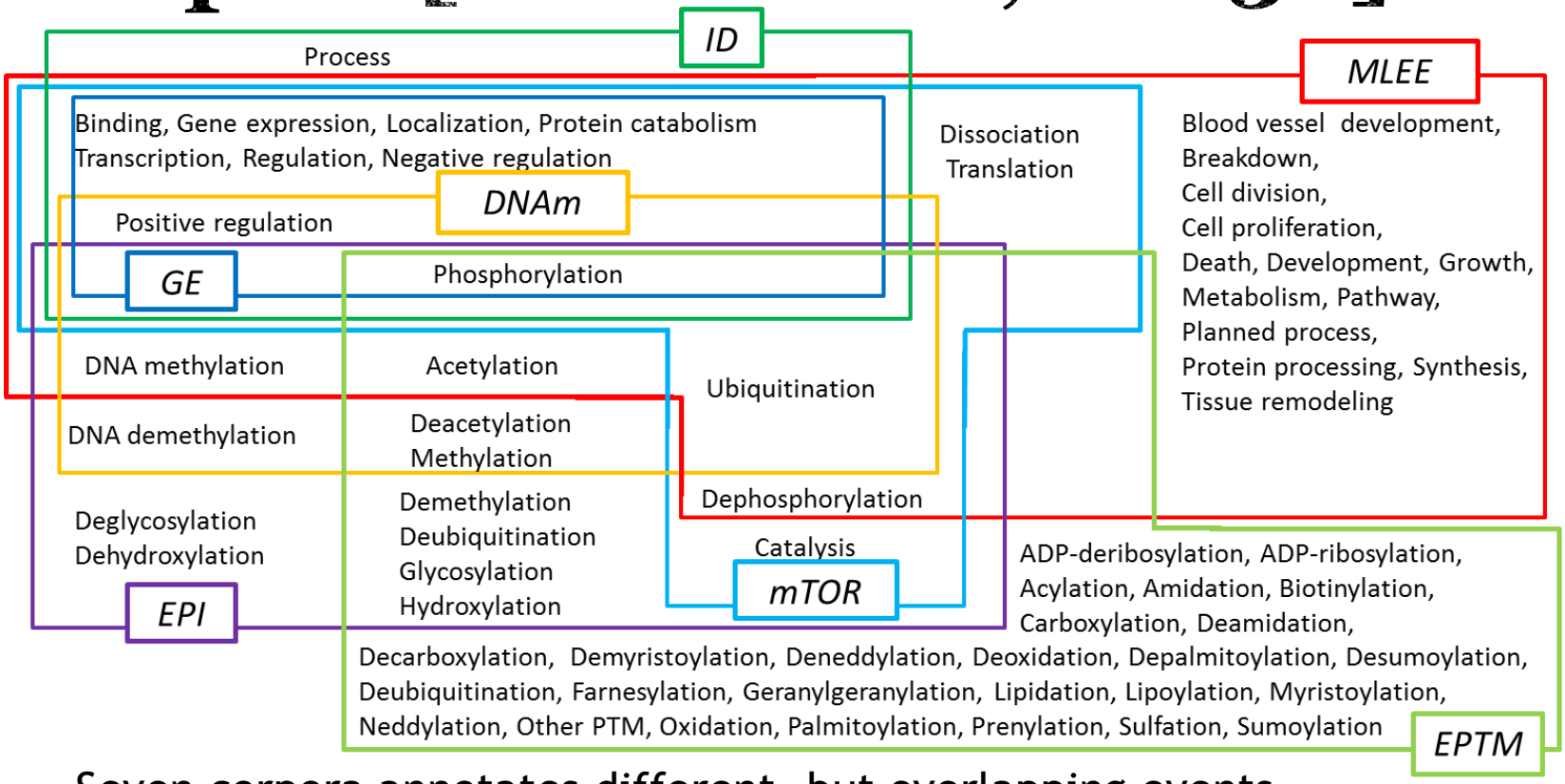
- Should we consider *bias-variance trade-off* and select features? ⇔ We do not know representative features, so an IE model with manually-selected features can easily underfit to training data
- Regularization
  - Regularization can well adjust the trade-off
  - AkaneRE and EventMine almost overfit the training data (close to 100% accuracy)
- Feature hashing
  - Memory cost reduction with slight performance loss
    - In EventMine, feature space (with hundreds of millions features) are condensed by **hashing** to 2^{20} (~1 million)

Addressing few instance problems

- Multi-task learning/domain adaptation/transfer learning
  - Several *different* corpora for PPIs and events [Miwa et al., 2009,Miwa et al., 2013b]
    - Different in definitions of types, domains, etc
  - Incorporating distant supervision, e.g., domain database
- Semi-supervised learning
  - Self-training/bootstrapping
  - Incorporating unsupervised information via word embeddings [Li et al., 2015]
    - Note: direct incorporation of word embeddings can hurt the performance of traditional systems [Guo et al., 2014]
- Adding domain features from thesaurus

  ⇔We hide entity names in biomedical IE, so we have no way to use them for entities.

Event extraction from partially overlapping corpora[Miwa et al., 2013b]



Seven corpora annotates different, but overlapping events

Event extraction from partially overlapping corpora[Miwa et al., 2013b]

- Method
  - Create training sets by unifying training instances in several corpora with removing unreliable examples
  - Train a single model on all the training sets
- Results
  - A model trained on multiple corpora outperforms all the individual models

	GENIA	ID	EPI	MLEE	DNAm	EPTM	mTOR
Individual	56.28	57.69	48.68	52.11	72.4	44.0	47.1
Multiple	57.28	59.06	54.35	52.76	76.0	50.0	51.0

Pipeline-specific problems

- Jack-knifing (cross-validation) on the training set
  - A system trained on gold annotations is biased and does not work well on predicted annotations.
    - Since the system never see wrong annotations, or it needs to fit to the annotations that are hard to be predicted

  ➔ Using prediction results by jack-knifing

- Tuning strategy of pipeline modules
  - Missing annotations in earlier stages cannot be recovered in the latter stages
  - Keeping too many unreliable instances, however, causes highly imbalanced problems

Incorporating deep learning (1)

- Convolutional or recurrent neural networks (CNN or RNN) methods often produce the best performance in several tasks
  - SemEval-2010 Task 8
    - Many DL-based models for relation classification
    - Best shared task system 82.2% (SVM) ➔ 88.0% (CNN) [Wang et al., 2016]
  - AIMed (PPI)
    - AkaneRE 65.2% (SVM) ➔ 72.4% (CNN) [Quan et al., 2016]
  - DDI Extraction 2013
    - 67.0% (SVM) [Kim et al., 2015] ➔ 70.5% (CNN) [Quan et al., 2016]
  - ACE Event
    - 52.7% (Structured perceptron) [Li et al., 2014] ➔ 55.4% (RNN) [Nguyen et al., 2016]

Incorporating deep learning (2)

- High performance often without external resources
  - We don't know what made this performance boosting, and we may get better with external resources
- Semi-supervised learning by incorporating pre-trained embeddings
  - Unlike feature-based learning, no preprocessing on embeddings is required
- Easy to build parameter-sharing models
  - Recent deep learning frameworks are quite helpful
    - Dynet, Chainer, Tensorflow, Theano, Keras, etc.

Incorporating deep learning (3)

- High computational costs
  - Can we apply the model to PubMed?
- High tuning costs
  - Parameters, model architectures, etc.
  - Performance is not stable due to non-convexity, parallelization, etc., and most results are not completely reproducible
  - The performance highly depends on tuning, and the comparison is quite difficult. (the models are good? Or the tuning is good?)
- Almost impossible to analyze what is learned
  - Feature engineering is less magical than model engineering

Minor points in building IE systems (1)

- Mismatches in automated preprocessing
  - Entities do not match linguistic units like words, phrases, or sentences
  - Relations can be inter-sentential due to errors in sentence splitting
- Selection of preprocessing modules including tokenizers, sentence splitters, parsers, stemmers, and named entity recognizers, etc.
  - Parser comparison on event extraction [Miwa et al., 2010] shows that the parsers and formats affect a lot (> 5 percent points) in F-score

Minor points in building IE systems (2)

- Data formats and conversions
  - byte offsets (old ascii systems) vs character offsets (recent UTF-8 systems)
  - XML (GENIA, Bioinfer, DDI, etc.) vs standoff (brat)
- Configurability and usability
  - Applicable to other tasks without changing the codes

Summary

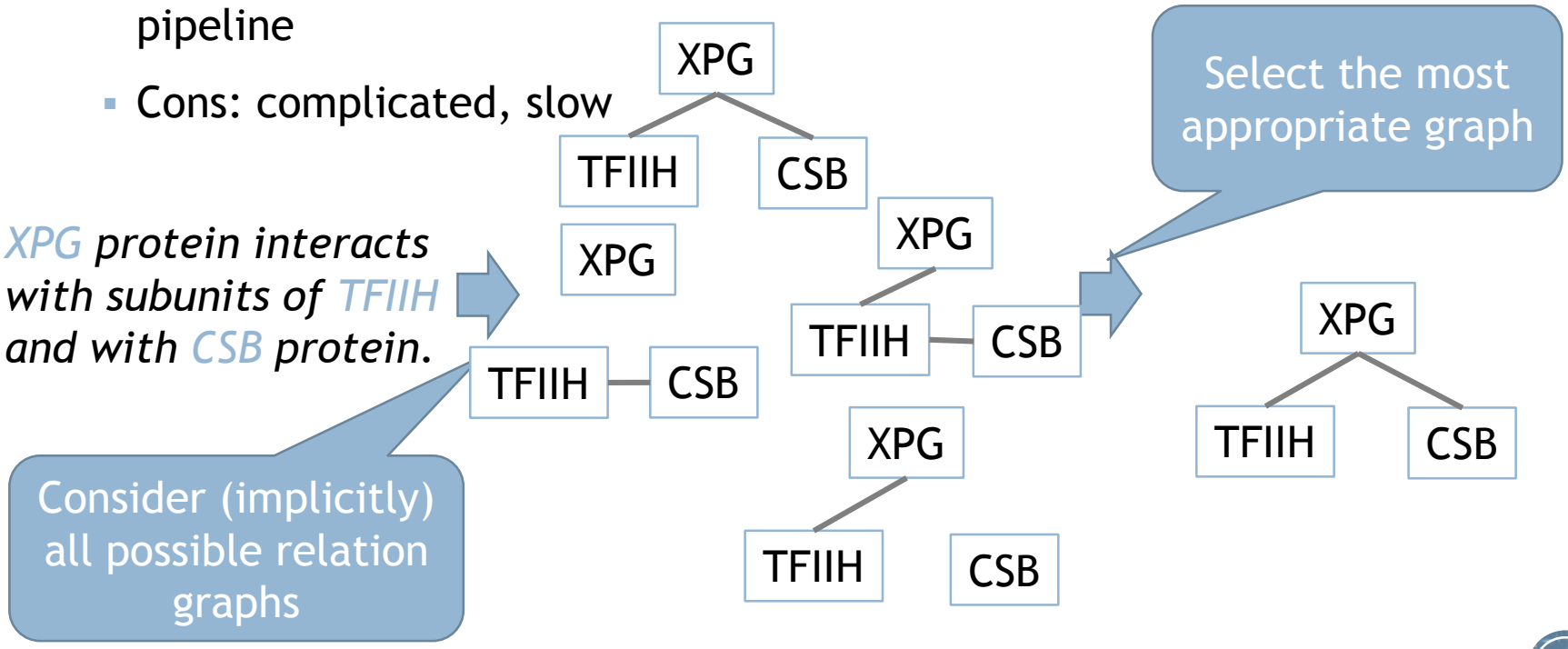
- In developing ML-based systems, we need to solve many problems that seem not to be essential for NLP/TM
  - Their handlings affect the performance in a non-negligible way

- Discussion
  - How can we evaluate each individual method?
  - What is solved and what is remaining problem?
    - Analysis on the state-of-the-art results is required

Joint entity and relation extraction

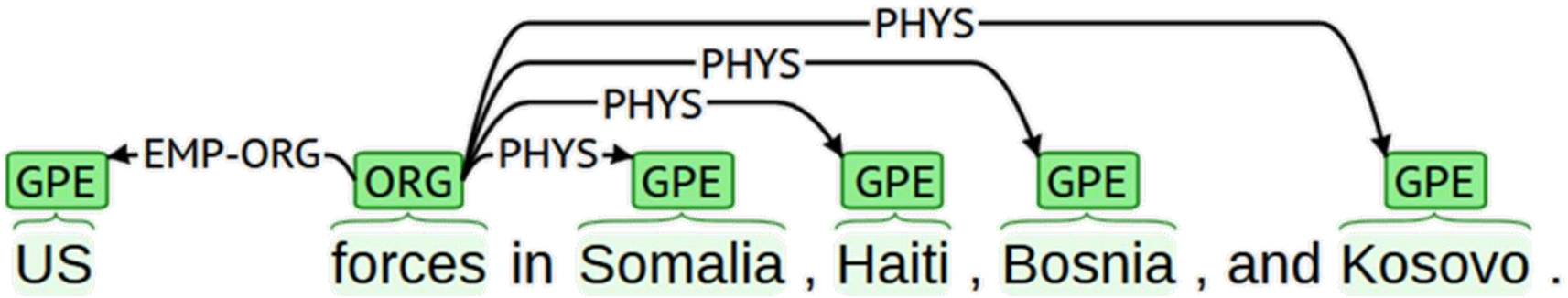
Joint entity and relation extraction

- Extract a relation graph from a target sentence
  - Pros: treatment of interactions among entities and relations, no pipeline
  - Cons: complicated, slow



Joint entity and relation extraction [Miwa et al., 2014b]

- Search-based structured prediction with global features
  - Features among multiple relations
    - E.g., “Somalia” and “Haiti” are in coordination, and they have “PHYS” relations with “forces”

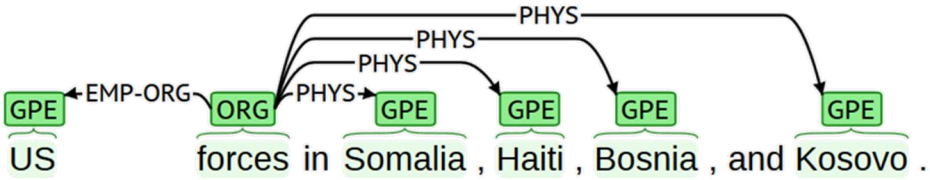


[Li et al., 2014] proposed similar approach

Joint entity and relation extraction

- Joint representation by table representations

➔Filling table cells one by one with search-based structured learning



	US	forces	in	Somalia	Haiti	Bosnia	and	Kosovo
US	GPE							
forces	EMP-ORG	ORG						
in	⊥	⊥	⊥					
Somalia	⊥	<PHYS	⊥	GPE				
Haiti	⊥	<PHYS	⊥	⊥	GPE			
Bosnia	⊥	<PHYS	⊥	⊥	⊥	GPE		
and	⊥	⊥	⊥	⊥	⊥	⊥	⊥	
Kosovo	⊥	<PHYS	⊥	⊥	⊥	⊥	⊥	GPE

Joint entity and relation extraction

- Joint learning performs well on relation extraction, but not entity
  - Entity detection performance dropped since the performance was tuned for relation extraction performance

➔The joint system may find more relation-related entities than the pipeline does

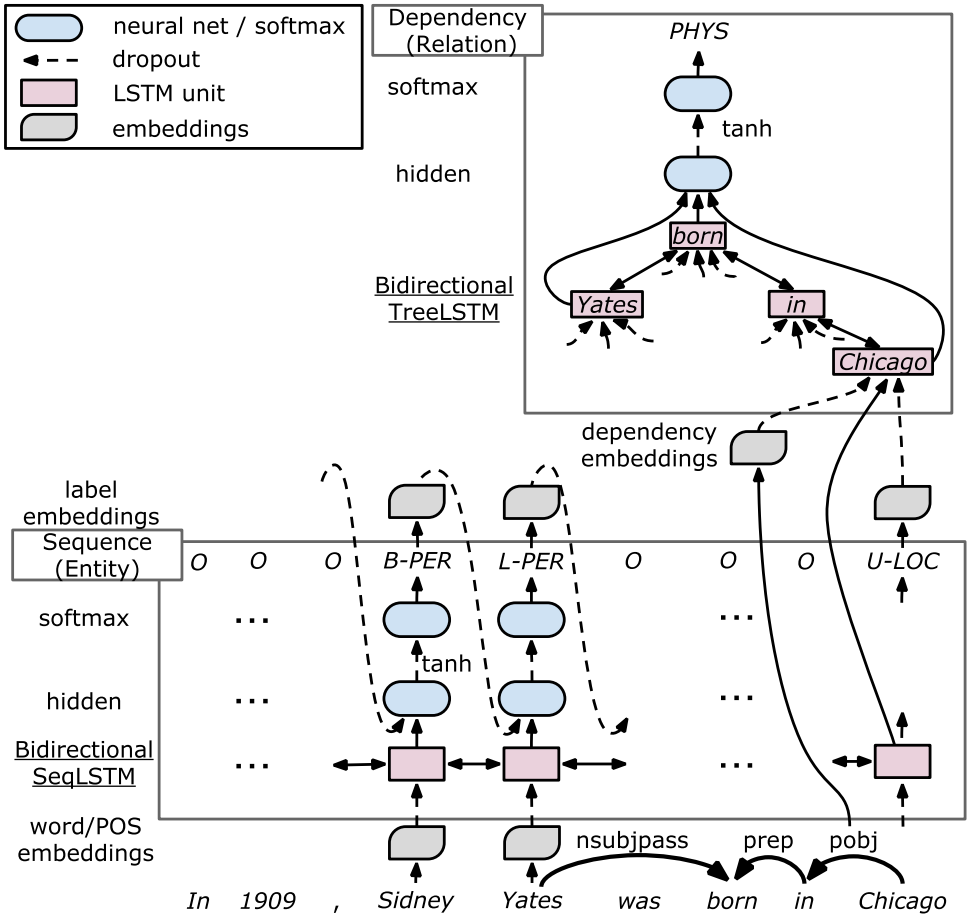
	Entity	Relation
Pipeline	81.8	57.7
Joint	81.3	61.2

F1-scores on CoNLL04 (%)

Available at https://github.com/tticoin/JointER

Deep learning-based entity and relation extraction [Miwa et al., 2016]

- Stacked RNN-based approach
  - Entity sequential LSTM-RNNs on word sequence
  - Relation tree LSTM-RNNs on parse tree
- Not structured prediction, but end-to-end approach
  - No structured margin loss, no beam search
  - Shared parameters among prediction tasks



Deep learning-based entity and relation extraction

- Deep learning-based greedy model outperforms feature-based structured learning model in a significant margin
- Deep learning can push forward the state-of-the-art, but the performance is still not high enough

	Entity	Relation
Miwa et al., 2016	83.4	55.6
Li et al., 2014	80.8	49.5

F1-scores on ACL2005 Relation (%)

Available at
https://github.com/tticoin/LSTM-ER

➔What can we import from traditional feature-based models?

Summary

- Joint entity & relation extraction models show better performance on relation extraction, and with deep learning, the performance was boosted
- Discussion
  - How can we analyze the model in detail?
  - Since deep learning models are flexible, we can incorporate external knowledge and information (e.g., database, distant supervision, multi-task learning)
    - We should relax the restriction on blinded named entities, since using word embeddings will keep generality

Conclusions

- I introduced our work from simple classification models to deep learning models, along with several problems to consider
  - We need to find a solid way to analyze errors

- Please try our models at our github:

  https://github.com/tticoin/

References (1)

[Miwa et al., 2016] Miwa, M, Bansal M. End-to-End Relation Extraction using LSTMs on Sequences and Tree Structures. In the Proceedings of ACL 2016. pp. 1105-1116.

[Black et al., 2016] Black, W. et al., 2016. Text Mining for Semantic Search in Europe PubMed Central Labs. E. Tonkin G. Tourte Eds, Working with Text. Chandos, pp. 111-131.

[Nhung et al., 2015] Nhung T. H. et al., 2015. Identifying synonymy between relational phrases using word embeddings. Journal of Biomedical Informatics, Vol. 56, pp. 94--102.

[Miwa et al., 2014a] Miwa, M., Thompson, P., Korkontzelos, I. Ananiadou, S., 2014. Comparable Study of Event Extraction in Newswire and Biomedical Domains. Dublin, Dublin City University and ACL, pp. 2270-2279.

[Miwa et al., 2014b] Miwa, M, Sasaki, Y. 2014. Modeling Joint Entity and Relation Extraction with Table Representation. In the Proceedings of EMNLP 2014. pp. 1858—1869.

[Miwa et al., 2013a] Miwa, M. et al., 2013. A method for integrating and ranking the evidence for biochemical pathways by mining reactions from text. Bioinformatics, Vol. 29, pp. i44--i52.

References (2)

[Miwa et al., 2013b] Miwa, M., Pyysalo, S., Ohta, T. Ananiadou, S., 2013. Wide coverage biomedical event extraction using multiple partially overlapping corpora. BMC Bioinformatics, Vol. 14, p. 175.

[Miwa et al., 2012] Miwa, M., Thompson, P. Ananiadou, S., 2012. Boosting automatic event extraction from the literature using domain adaptation and coreference resolution. Bioinformatics, Vol. 28, pp. 1759-1765.

[Miwa et al., 2010] Miwa, M., Pyysalo, S., Hara, T. Tsujii, J., 2010. Evaluating Dependency Representations for Event Extraction. In the Proceedings of COLING 2010, pp. 779-787.

[Miwa et al., 2009] Miwa, M., Sætre, R., Miyao, Y. Tsujii, J., 2009. A Rich Feature Vector for Protein-Protein Interaction Extraction from Multiple Corpora. In the Proceedings of ACL-IJCNLP2009, pp. 121-130.

[Quan et al., 2016] Quan, C., Hua, L., Sun, X. Bai, W., 2016. Multi-channel convolutional neural network for biological relation extraction. BioMed Research International. Vol. 2016.

[Wang et al., 2016] Wang, L., Cao, Z., de Melo, G. Liu, Z., 2016. Relation Classification via Multi-Level Attention CNNs. In the Proceedings of ACL2016, pp. 1298-1307.

References (3)

[Nguyen et al., 2016] Nguyen, T. H., Cho, K. Grishman, R., 2016. Joint Event Extraction via Recurrent Neural Networks. In the Proceedings of NAACL2016, pp. 300-309.

[Kim et al., 2015] Kim, S., Liu, H., Yeganova, L., Wilbur J. 2015. Extracting drug–drug interactions from literature using a rich feature-based linear kernel approach. Journal of Biomedical Informatics, vol. 55, pp. 23–30.

[Guo et al., 2014] Guo, J., Che, W., Wang, H., Liu, T. Revisiting Embedding Features for Simple Semi-supervised Learning, In the Proceedings of EMNLP2014, pp. 110-120.

[Li et al., 2014] Li, Q., Ji, H., Huang, L., 2013. Joint Event Extraction via Structured Prediction with Global Features, In the Proceedings of ACL2014, pp. 73-82.

[Li et al., 2013] Li, Q., Ji, H, 2014. Incremental Joint Extraction of Entity Mentions and Relations. In the Proceedings of ACL2013, pp. 402-412.

[Stenetorp et al., 2012] Stenetorp, P. et al., 2012. BRAT: a web-based tool for NLP-assisted text annotation. In the Proceedings of EACL2012, pp. 102-107.