

# 生医学文書を対象とした情報抽出タスク

東京大学大学院情報理工学系研究科 三輪 誠, 辻井潤一

タスク定義		BioNLP Shared Task 2009*1	BioCreative 2.5*2
開発期間		2008年12月-2009年3月	2009年2月-2009年6月
抽出対象	対象	イベント (Task 1)	関係 (IPT task)
	関係	1対多, 再帰構造	1対1
	方向	有	無
	イベントクラス	9	無
	関係クラス	2	無
データ	学習	800 アブストラクト	740 論文
	評価	アノテーションレベル	論文
出力	入力	アブストラクト	論文
	蛋白質	既知	未知
	蛋白質	-	辞書ID
	対象	イベント	IDのペア
評価	極性	全て	Positive interactionのみ
	曖昧性	全て	断定のみ
	証拠	不必要	必要
評価	基準	F-score (正解/不正解)	AUC-iP/R (ランキング)
	評価対象	データ	サーバ(online) データ(offline)
投稿可能な結果		1	5x2

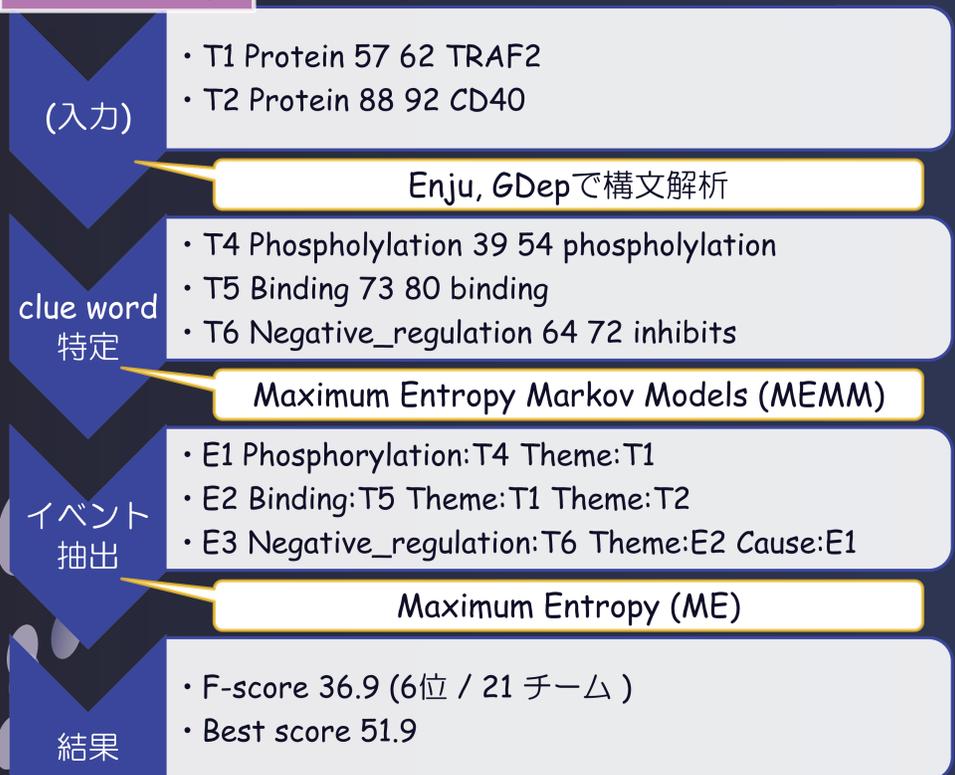
\*1 <http://www-tsujii.is.s.u-tokyo.ac.jp/GENIA/SharedTask/>  
 \*2 <http://www.biocreative.org/>

## BioNLP Shared Task 2009

## BioNLP'09 Shared Task on Event Extraction

... In this study we hypothesized that the phosphorylation of TRAF2 inhibits binding to the CD40 cytoplasmic domain. ...

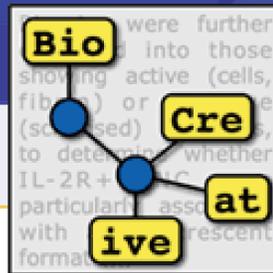
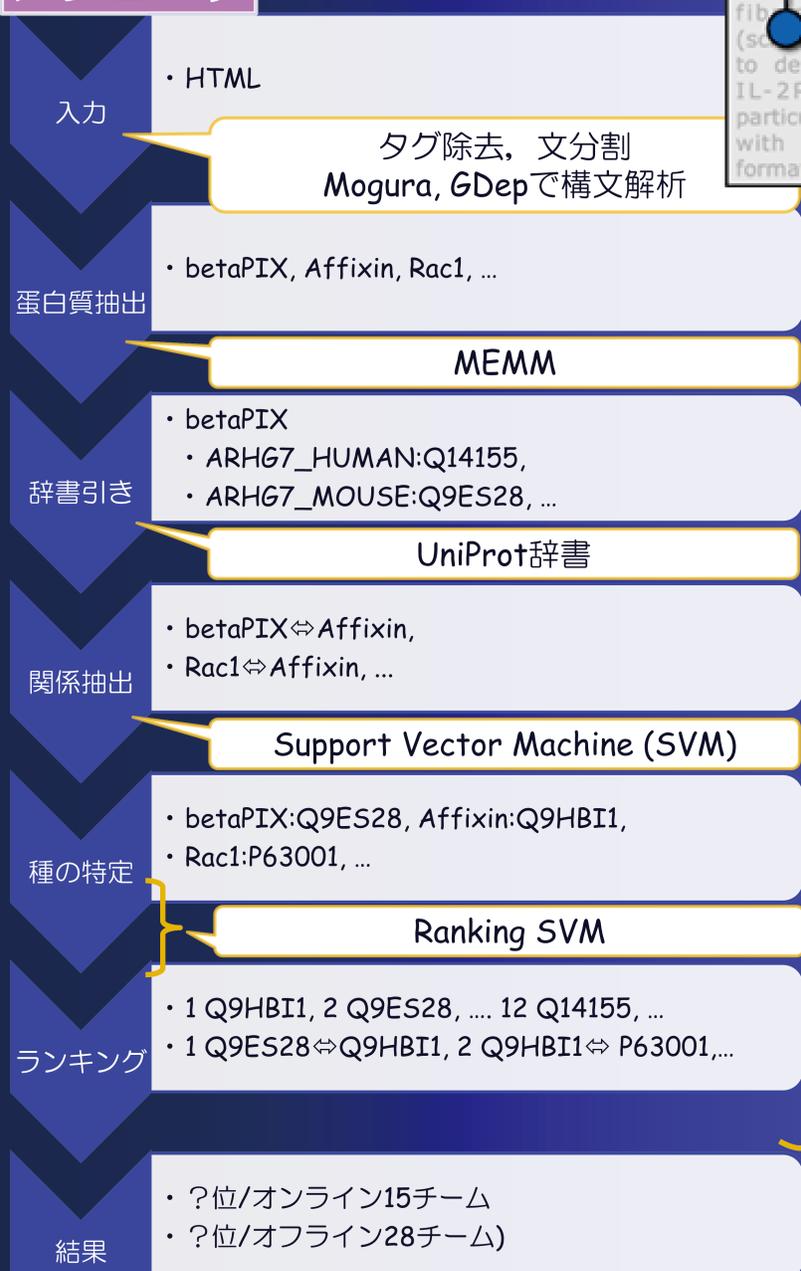
### アプローチ



## BioCreative 2.5

**SUMMARY:** Immunofluorescence analysis of stable C2C12 cells expressing human affixin...  
**TITLE:** Affixin activates Rac1 via betaPIX in ...  
**ABSTRACT:** Affixin/beta-parvin is an ...  
**KEYWORD:** p21-activated kinase , ..., betaPIX, ...  
**BODY:** Affixin/beta-parvin (affixin) [1,2] is one of family of parvin family together with alpha-parvin/actopaxin/CH-ILKBP... The C2C12 myoblast is derived from mouse satellite cell ...

### アプローチ



### まとめ (+a)

- 2つの関係抽出タスクについて紹介
  - オンラインシステムは公開中 (BC2.5)
  - 上位システムは公開予定 (ST)
  - 両タスクともデータは利用可能
- BioNLP 2009 workshopは6月開催済, BC2.5 workshopは10月開催予定 (結果発表)
- まだ, 始まったばかり
  - STで現在F-score 53.2

- UIMA ワークフロー (U-Compare)
- Webサービス(online)

# 生医学文書を対象とした 情報抽出タスク

東京大学大学院

三輪 誠 辻井潤一

# 発表の概要

- 単純な1文中の2項関係抽出から発展して定義された2つの関係抽出に関するShared Task
  - BioNLP 2009 Shared Task
  - BioCreative 2.5
- 生医学の分野において重要なタスク
  - 想定されるアプリケーション
    - 蛋白質の関係グラフ (パスウェイ) 構築
    - 論文データベースの自動作成

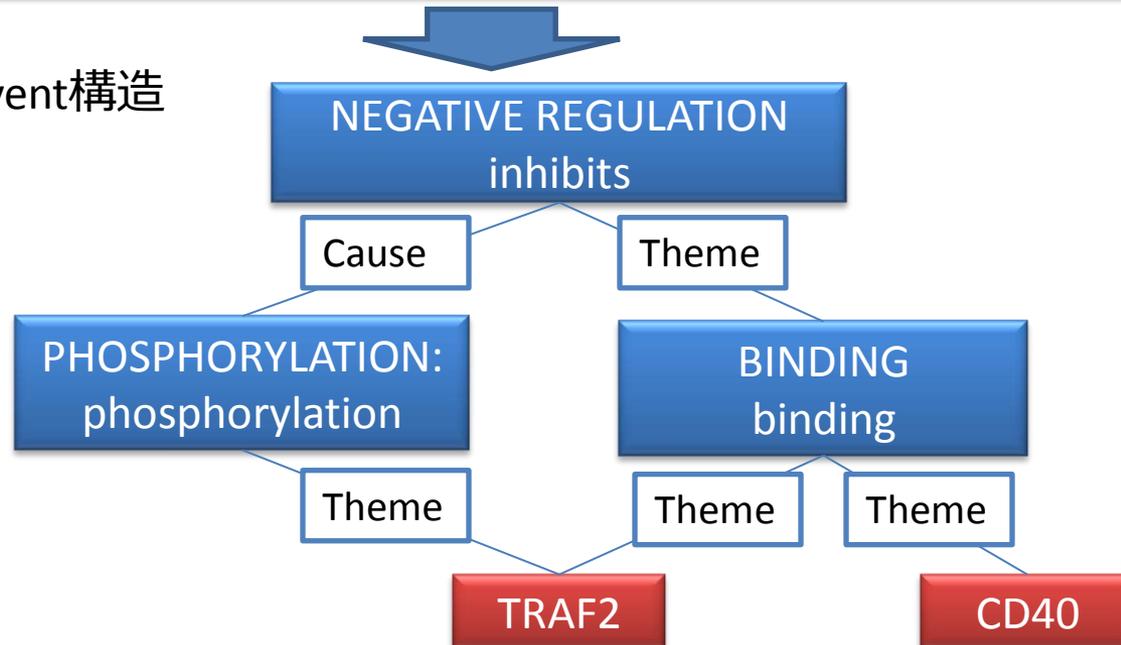
# BioNLP'09 Shared Task

- 文書中の既知の蛋白質間のイベント構造の抽出

入力:論文のアブストラクト, 蛋白質 (固有表現)

... In this study we hypothesized that the phosphorylation of *TRAF2* inhibits binding to the *CD40* cytoplasmic domain. ...

出力: Event構造



# BioCreative 2.5

- HTML化された論文から蛋白質 (の辞書ID) 間の相互作用抽出

入力: 論文 (HTML)

```
<SUMMARY>Immunofluorescence analysis of stable C2C12 cells expressing human affixin....</SUMMARY>  
<TITLE> Affixin activates Rac1 via betaPIX in ... </TITLE>  
<ABSTRACT> Affixin/beta-parvin is an ... </ABSTRACT>  
<KEYWORD> p21-activated kinase , ..., betaPIX, ... </KEYWORD>  
<BODY> Affixin/beta-parvin (affixin) [1,2] is one of family of parvin family together with alpha-parvin/actopaxin/CH-ILKBP... The C2C12 myo-blast is derived from mouse satellite cell ...<BODY>
```



出力: 相互作用している蛋白質(辞書ID)のペア

•Q9ES28 ⇔ Q9HBI1

•betaPIX ⇔ Affixin

•ARHG7\_MOUSE ⇔ PARVB\_HUMAN

•Q9HBI1 ⇔ P63001

•Affixin ⇔ Rac1

•PARVB\_HUMAN ⇔ RAC1\_MOUSE

...

# ポスターでは

- タスク設定紹介, 我々のシステムの概要紹介, 参加報告
  - BioNLP 2009 Shared Task
    - 2008年12月～2009年3月
  - BioCreative 2.5
    - 2009年2月～2009年6月