

# Entity Linking

- 目的
  - ▶ 表記揺れや曖昧性の問題を解決しながら、発言の根拠となる一次情報と結びつける
  - ▶ PolInfo2 では「法律名」を対象とする

- 入力
  - ▶ 出題ファイル（形態素）
  - ▶ Wikipedia（2019-12-01）

- 出力
  - ▶ 出題ファイル（形態素, IOB2, メンション, URL）

- ✓ メンション抽出結果：IOB2 タグ
- ✓ 曖昧性解消結果：メンションに対応するWikipediaのタイトル

形態素	IOB2	メンション	Wikipedia URL
この			
議場			
で			
過労	B	過労死促進法過労死等防止対策推進法過労死促進法	<a href="http://...">http://...</a>
死	I	過労死促進法過労死等防止対策推進法過労死促進法	<a href="http://...">http://...</a>
等	I	過労死促進法過労死等防止対策推進法過労死促進法	<a href="http://...">http://...</a>
防止	I	過労死促進法過労死等防止対策推進法過労死促進法	<a href="http://...">http://...</a>
対策	I	過労死促進法過労死等防止対策推進法過労死促進法	<a href="http://...">http://...</a>
推進	I	過労死促進法過労死等防止対策推進法過労死促進法	<a href="http://...">http://...</a>
法	I	過労死促進法過労死等防止対策推進法過労死促進法	<a href="http://...">http://...</a>
が			

- 評価
  - ▶ 適合率, 再現率, F値

データセット	ファイル数	形態素数	ファイルサイズ合計
学習用データ	1	260,366	2.7MB
評価用データ	1	209,862	1.9MB

# Entity Linking の3つの課題

1. 表記ゆれの解決
2. 曖昧性の解決
3. リンク先が存在するのか、存在しないか、明確にする

・特定複合観光施設区域整備法案、いわゆるIR整備法案について、最近の世論調査では、カジノ法案の成立は不要としている国民の方々七六%、自民党の支持の方々でも六四%に及びます。

正式名称

特定複合観光施設区域整備法

Wikipediaに存在しない

・また、一昨年のIR推進法の審議の際には……

・こうした問題点を踏まえれば、IR法案は、我が国には必要のないざる法、悪法と言わざるを得ません。

特定複合観光施設区域の整備の推進に関する法律

・その当時のカジノ法案、カジノ解禁法案、IR法案があったんですよね。

Wikipediaに存在する

# メンション抽出

テキスト中のメンションを抽出するタスク

メンションの表現にはIOB2タグがよく使われる

入力

一 昨 年 の IR 推 進 法 の 審 議 の 際 に は 、 …

出力

一 昨 年 の IR 推 進 法 の 審 議 の 際 に は 、 …  
O O O B I I O O O O O O …

B(Begin) : メンションの始まりを示す

I(Inside) : メンションが続いていることを示す

O(Outside) : メンション以外を示す

# メンションの曖昧性解消

抽出したメンションに対し、知識ベースから候補エンティティを生成  
候補エンティティの中から尤もらしいエンティティを結びつける

入力

一	昨	年	の	IR	推	進	法	の	審	議	の	際	に	は	、	...
0	0	0		B	I	I		0	0	0	0	0	0	0	0	...

候補  
エンティティ

特定複合観光施設区域の整備の推進に関する法律
建設工事従事者の安全及び健康の確保の推進に関する法律

出力

特定複合観光施設区域の整備の推進に関する法律