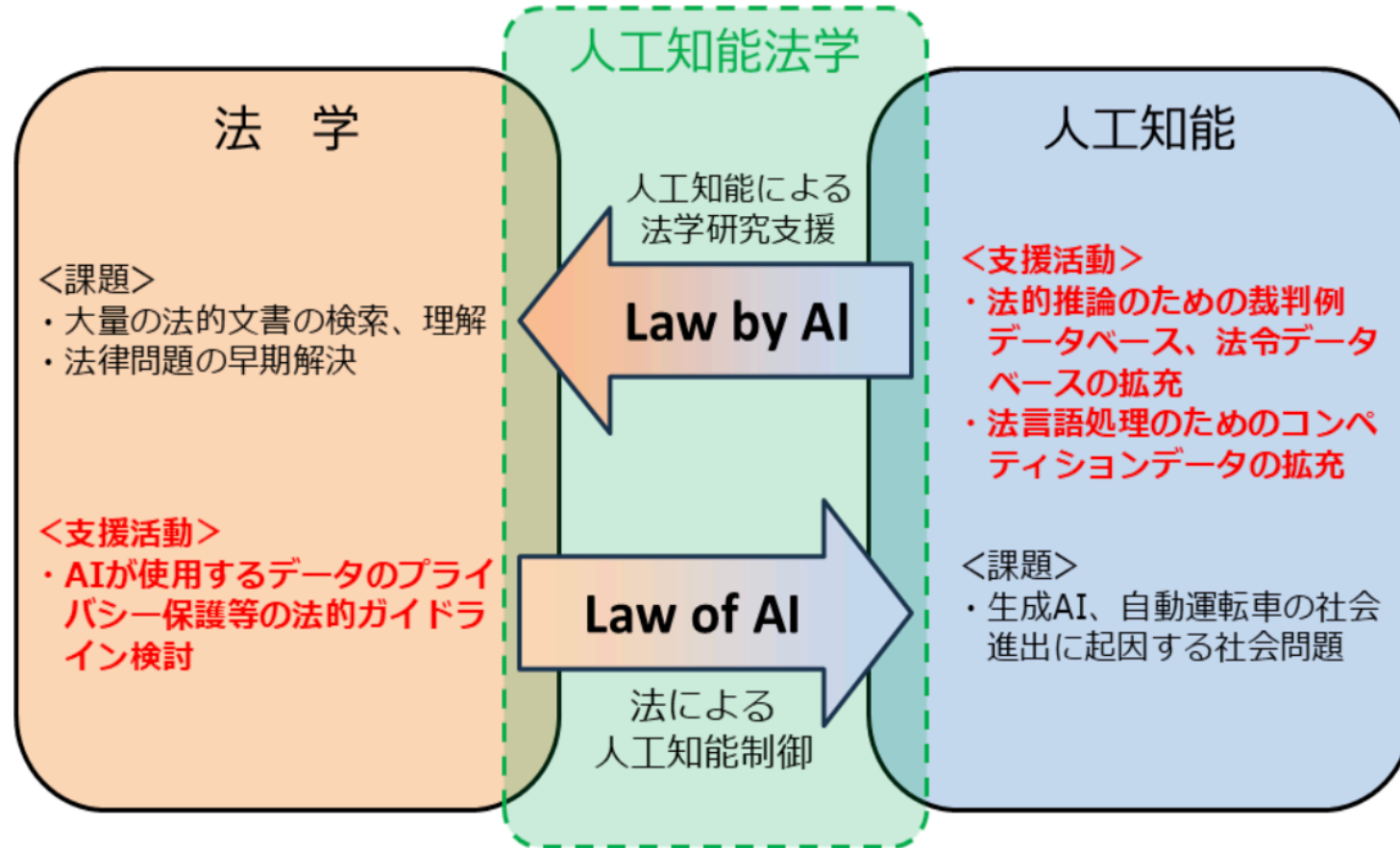




日本の判例データ

情報・システム研究機構
データサイエンス共同利用基盤施設
人工知能法学研究センター
西野文人

人工知能法学研究センターの活動



判例データの活用支援

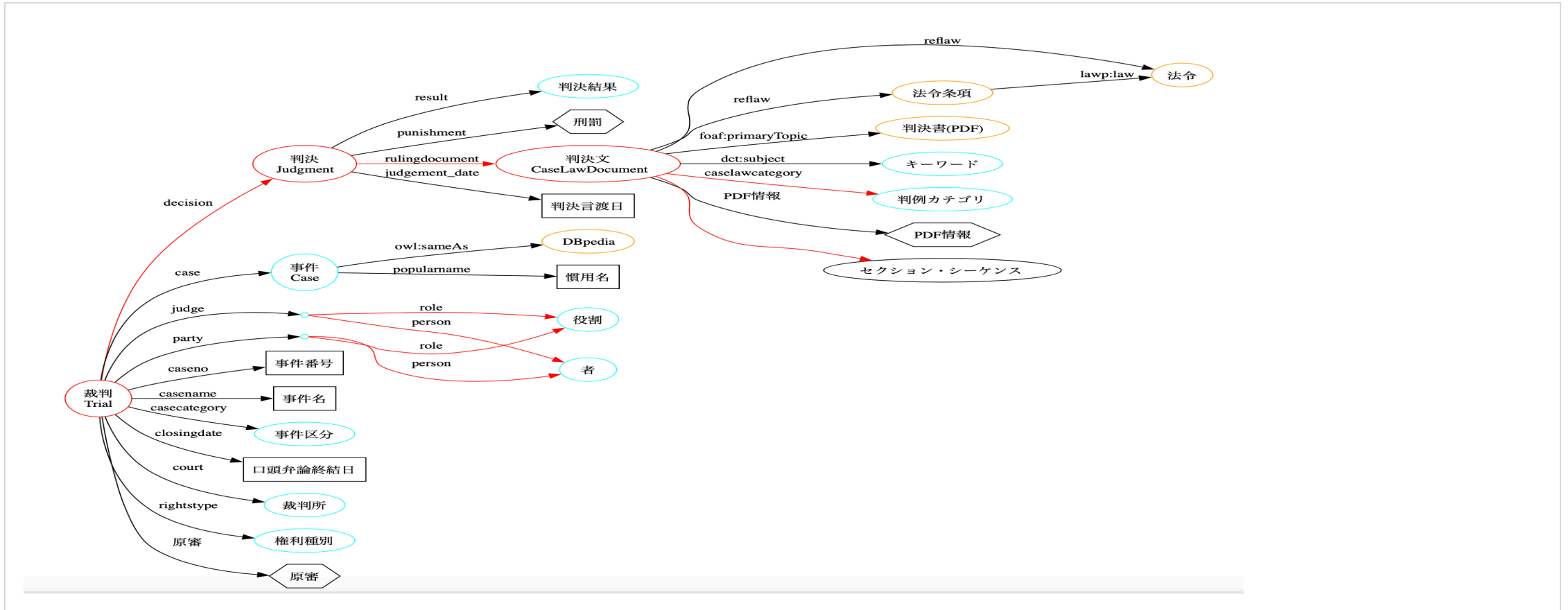
- 背景：判例データは最高裁判所サイトで検索・閲覧できるが、PDFでの提供、一括ダウンロードなし
→ データとして扱えない
- 対応：
 - 自然言語処理用データとして扱えるようにする → PDFの収集・HTML化
 - 判例の高度な検索・分析を行えるようにする → ナレッジグラフ (Linked Data) 化
- 応用例：判例ナレッジグラフのフロントエンド

論理構造HTML化処理

1. 判例のメタデータの解析、PDFのダウンロード（検索、検索結果の解析、PDFのダウンロード）
2. PDFの解析（ワープロ特有のPDFの処理など）
3. 文書論理構造の解析（不要部の削除、小見出し・箇条書き・段落認識など）
4. 判決書構造（目次,表題部,主文,事実及び理由,末尾事項,附属書類）を認識【新規】
→ 構造認識誤りの削減、意味構造に応じた処理が可能

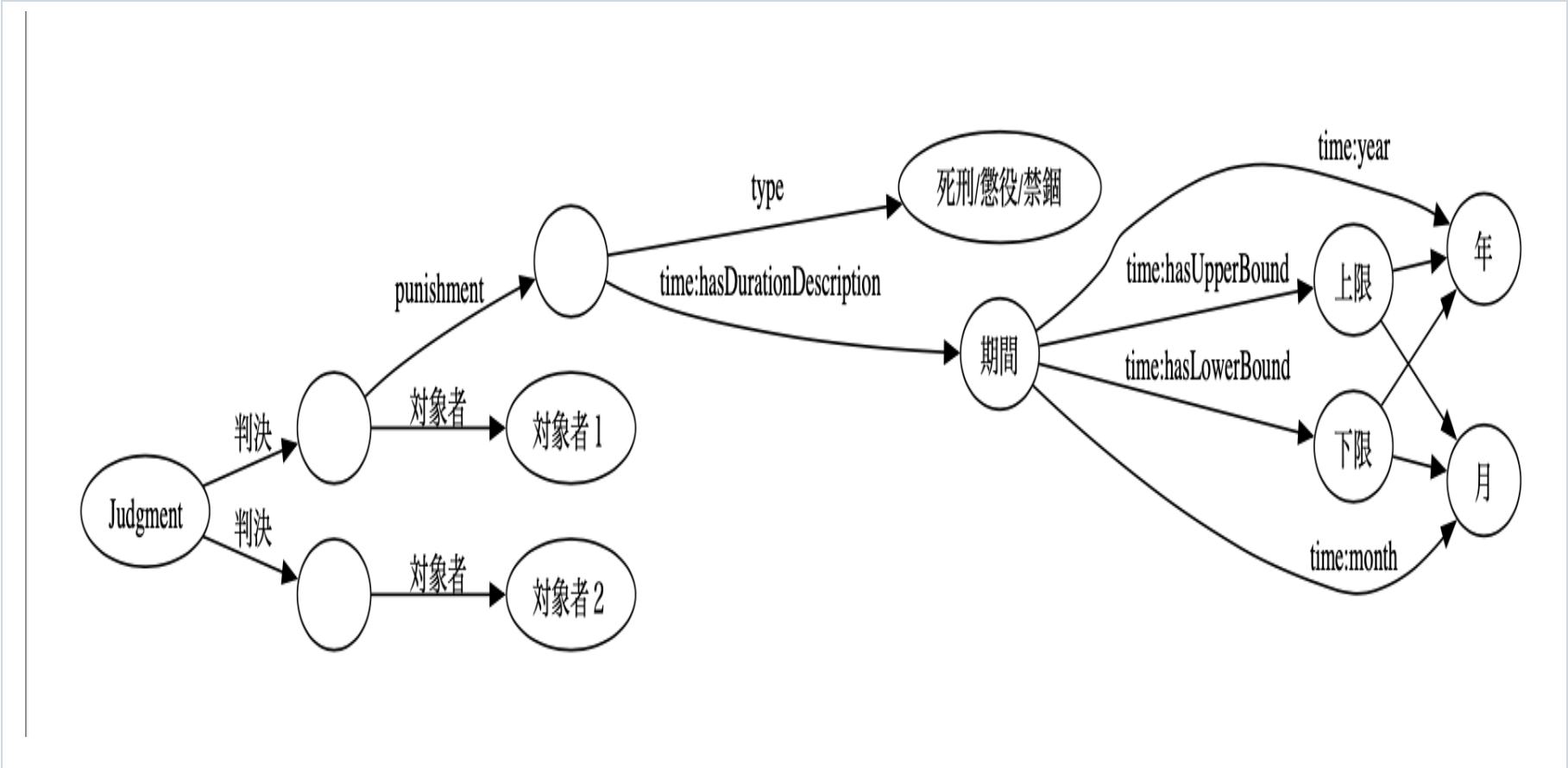
日本の判例オントロジーの作成

主要部構成図



日本の判例オントロジーの作成

刑罰部



各種情報の抽出とRDF化

- 最高裁判所メタデータ（事件名、裁判年月日、法廷名、裁判種別、結果、原審事件番号、判示事項, 裁判要旨など）からのRDF化【強化】
- テキストからの抽出（裁判官、当事者、参照法令、キーワード、参照特許【新規】、刑罰【新規】、判決書構造【新規】など）
- PDFやHTMLから取得した文書に関する情報（PDFページ数、divタグやbrタグの数）【新規】
- 外部データベースへのリンク(DBpedia, eLaws, GooglePatents【新規】)
- 現在の判例RDF：約2000万triples

KGナビゲータ

- SPARQLを使わずにRDFを閲覧するツール（データ駆動ダイナミックパブリッシング）
 - 焦点エンティティの関連情報をSPARQL検索で抽出す
 - 指定された形式（表やチャートなど）で表示する
 - ページ構成、検索式、表示形式はyamlで記述

KGナビゲータ画面例

宮川美津子(Person)

Involved

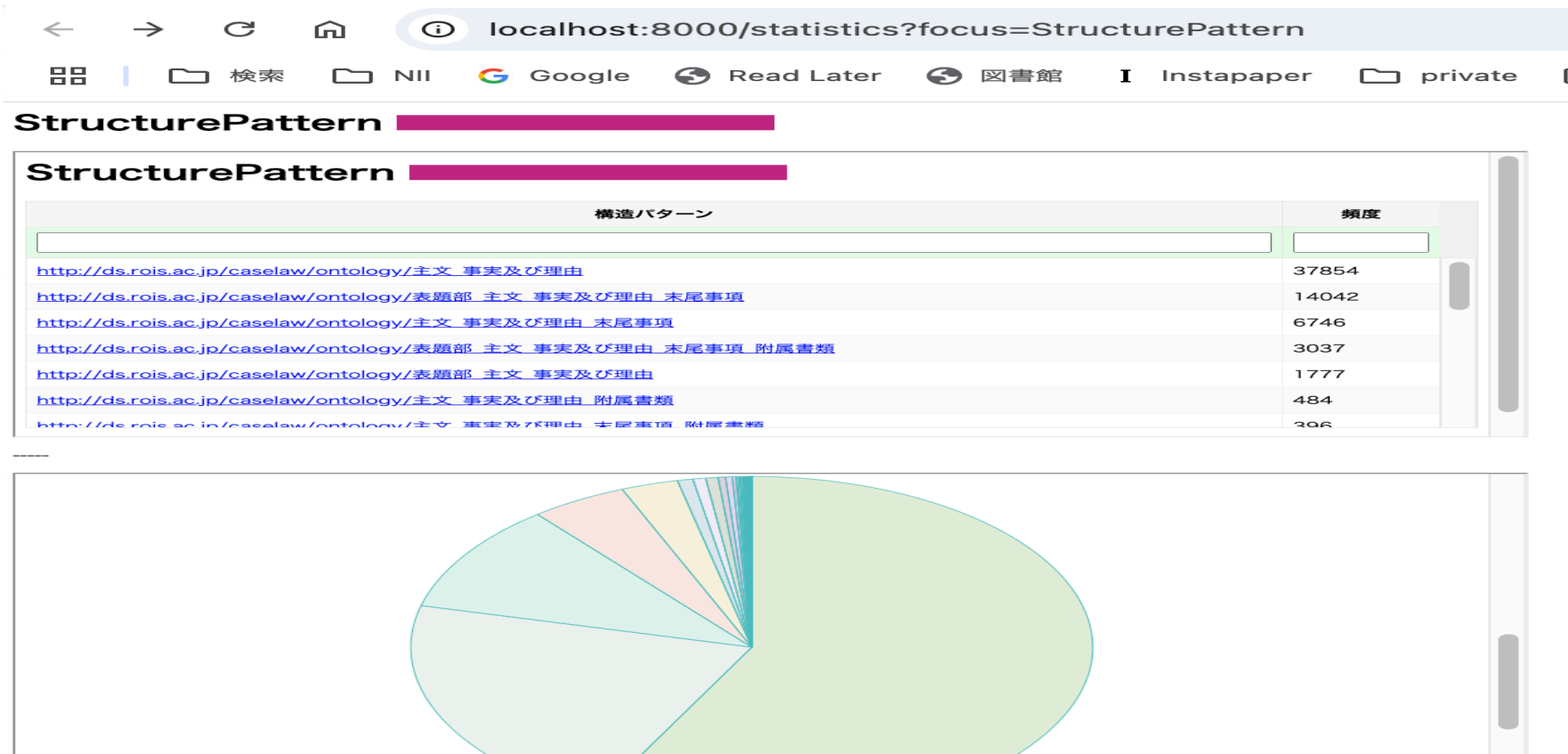
関与した裁判 (事件番号)	関与した裁判 (事件名)	役割
令和4ネ10051	不正競争行為差止等請求控訴事件	訴訟代理人弁護士
令和4行ケ10089	審決取消請求事件	訴訟代理人弁護士
令和5さ24	道路交通法違反被告事件に係る略式命令に対する非常上告事件	裁判官
令和5あ292	窃盗、道路交通法違反、殺人被告事件	裁判官
令和4受1744	賃料減額等請求事件	裁判官
令和4行ヒ319	懲戒処分等取消請求事件	裁判官
令和5受1210	国家賠償請求事件	裁判官

Profile

attribute	value
名称	
リンク	http://ja.dbpedia.org/resource/宮川美津子
説明	宮川 美津子（みやがわ みつこ、1960年2月13日 - ）は日本の弁護士、最所判事。

KGナビゲータ画面例（統計）

- 一つのリソースのURIではなく、「ページ数」「判決書構造」のような注目属性をもつURIに対してもページを定義



議論歓迎

- 判例データ (HTML、RDF) を活用したい人
- ナレッジグラフやナレッジグラフナビゲータに興味のある方
- PDF処理に興味
 - PDFで何が起こっているか
 - NotebookLM ではPDFをうまく処理できているのか
- 日本における判例、法令、特許などのデータ活用の問題に興味