

意味グラフのマッチングによる 事故問い合わせ文からの判例検索システム JCare

原田 実 * 鈴木 亮 **

青山学院大学 理工学部 情報テクノロジー学科 *

青山学院大学 理工学研究科 経営工学専攻 **

意味解析を用いた情報検索の一手法を提案し、「判例」を検索対象とし日本語文章で記述した「問い合わせ文」を検索質問とした検索システム JCare を開発する。本研究では文章が表す内容を、語が格納されたノードと語間の関係(深層格)を表すアークからなる意味グラフとして捉え、判例文と問い合わせ文の意味グラフ間における位相同型部分の大きさをもとに、文章間の内容類似度を算出する手法を提案する。このとき検索の高速化・精度向上の目的で View という考え方を提案する。文章を視点(View)により分類し、意味グラフを分割した View グラフの類似度を求めることで、内容的に関連性の低い文章間の計算時間、またそこから生まれるノイズを排除する。

Judicial CASE REtriever JCare from Japanese Query Sentences based on Semantic Graph Matching

Minoru Harada * Ryo Suzuki **

Department of Integrated Information Technology, Faculty of Science and Engineering,
Aoyama Gakuin University *

Graduate School of Industrial and System Engineering, Department of Science and Engineering,
Aoyama Gakuin University **

We propose the technique to realize the information retrieval based on the semantic analysis and actually develop the judicial case retrieval system called JCare. It accepts a query written in Japanese sentences and retrieves a judicial case containing the sentences describing the similar situation specified by the query. It first transforms both a case and the query into semantic graphs that have nodes representing the meaning of word and arcs representing the relations (deep case) between the words. Next, it calculates the similarity between the case and the query by searching the maximum common parts that are topologically equivalent. The graph matching is speeded up by separating each semantic graph into sub-graphs based on the "View" of a case.

1. はじめに

1.1. 研究背景

今日ある検索システムは、索引語を用いたキーワード検索が主流となっている。検索漏れを

防ぐために、「キーワードに指定した語の同意語や関連語も自動的に検索対象にする」といった工夫が凝らされているものも幾つか存在する。しかし、一般にキーワードによる絞込みは難しく、検索結果からまさに必要とする情報に絞り

込むには、その内容の説明文などを検索要求と比べる必要があった。もし検索システムが有能な秘書のように、必要な情報の説明を文章で与えるだけで検索対象の要約などの解説文の内容までを考慮して最適な情報を掲示してくれると、ユーザの検索労力は大幅に軽減される。

このような日本語文章を検索インターフェースに用いている研究には、例えば京都大学総合情報メディアセンターで公開されている Unix の利用方法に関するアドバイスシステムがある¹。このシステムは質問文の構文木と解説文の条件部の構文木を比較し、一致点に対して重みを付けて合計することによって類似度を求め、最も類似する解説文の結果部を表示するというものである。しかしここでは意味解析は行われておらず、したがって単純な質問文にしか答えられない。また的外れな解答もたまにあり、語の相互関係に敏感な質問には答えられない、などの問題点を有している。

このような質問文に対しては、意味解析や文脈解析を行った結果による意味レベルの類似性による情報検索が必要と考えられる。

1.2. 研究目的

本研究では、意味解析を用いた情報検索の一手法を提案する。具体的には「判例」を検索対象とし、自然言語で記述された「問い合わせ文」を検索質問とした判例検索システム JCare(Judicial CAse REtriever based on semantic graph matching)を開発する。本システムでは、自然語意味解析により「問い合わせ文」と「判例」の双方を意味グラフに展開し、意味的に同型な部分グラフを求めることで類似度を算出する。これにより両者の内容にまで踏み込んだ検索を実現する。

検索対象は「判例」の中でも「交通事故関連の判例」に絞り込む。「交通事故」の判例には“当事者”が存在し、それぞれの“当事者”が相互に「関係」を持つという特徴がある。”当

事者”とは事故に関係した人物をさし、被告、原告、被害者などがこれにあたる。この特徴により、照合時における比較基準が設定しやすくなる。また判例検索は社会的にも有用性が高い。これらの理由から判例を検索対象として採用した。

2. 意味グラフ

2.1. 検索対象－「判例」

「最高裁判所判例集」は最高裁判所判例委員会により選択されまとめられたもので、事件の判決を集めた「事例集」の中から特殊な事件とされるもの、特殊な判決（裁判上の先例となる判決）とされるものを集めたものである。現在約 50 万件に達する「事例集」に対して「判例集」には約 3000～4000 件が含まれている。この「判例集」に含まれるひとつひとつが「判例」である。

「判例」は大きく分けて、判例 ID、事件の種類、判決日、当事者の氏名などといった判例情報と、その判例の内容自体を記述した内容情報から構成されている。内容情報は約 8000 字で記述されている。この内容情報はさらに、事件の事実を表すもの、その事実から裁判官が下した判決とその理由に分けることができる。

判例の中に記述された事実文は大きく 2 つに分けることができる。ひとつは「当事者間に争いのない事実」を記述したもの、もうひとつは「当事者それぞれが主張する事実」を記述したものである。前者は当事者のもつ属性や事故のおきた場所など、確実に事実として認められるものを示す。これに対して後者は、確実に事実といえないものである。最終的に裁判官が、後者のものに対して確かな「事実」として認めるか否かの判断を下すのである。この時、裁判官は前者の「争いのない事実文」を事故当時の状況を判断するための要素として利用する。そこで、判決を下す上で事故の状況を示す「争い

のない事実文」が大きな役割を担っていると考えた。よって本研究では、検索対象を「当事者間に争いのない事実」文とする。

実際の判例検索システムでは、「判例」の中からこの「争いのない事実文」を検索対象として取り出す必要がある。この抽出プロセスは、本研究では手作業による前処理として実現する。その自動化は今後の課題とする。

2.2. 検索質問－「問い合わせ文」

本システムを利用するにあたって利用者は、図 1 に示すような、「事件の状況を示す事実を記述した文章」を「問い合わせ文」として与えるものとする。ここでは、事故当時の現場の状況、当事者の行動、当事者の属性などの、事故の発生に関する事実関係を明確に記述する。JCare はこの問い合わせ文に記述されている事故に意味的に類似した事故に関する判例を検索する。

本件事故現場は、南北に走る国道三五号線と環状八号線が交差する丁字路交差点である。この交差点において、被害者が運転する自動車に、被告が運転する自動車衝突し、被害者太郎が負傷した。
 本件事故現場には交通信号機が設置されている。路面はアスファルト舗装されており、乾燥していた。
 被告鈴木は、北方向から南方向に向かって時速約四〇キロメートルで進行していた。被告鈴木はその時に本件交差点の対面信号が黄色の点滅信号であることを認めたが、減速せずにそのまま進んだ。
 原告太郎は、被害車両に乗って、田園調布方面から環状八号線の第二車線を走行していた。しかし、本件交差点の少なくとも二つ手前の交差点を青信号により発進し、本件交差点の手前の交差点では停止することなく、本件交差点にさしかかった。

図 1 問い合わせ文

2.3. 意味解析（自然言語処理）

実際に検索を行うには、問い合わせ文と判例をその意味的な内容を表す内部表現に変換する必要がある。このためには言語表現からその意味を抽出するための一連の自然言語処理を行う必要がある。本研究ではこの過程に、形態素解析に奈良先端科学技術大学院大学の松本研究室が開発した茶筌²を、係り受け解析に同研究室が開発した茶掛を、意味解析に青山学院大学の原田研究室が開発した SAGE^{3,4}を、さらに文脈解析に同研究室が開発した InSeRA⁵を用いることとする。

```

frame述語
frame(1,南北,'ナンボク','JN1','南北'名詞一般,'none','10167',[],1).
frame(2,走る,'ハシル','JVB','走る'動詞-自立,'基本形','103ac',[],[gon,1],1).
...
frame(7,'国道三五号線','none','JPR','国道三五号線'
'none','none','000000',[],[consist,3],[consist,4],[consist,5],[consist,6],1).
...
frame(12,'環状八号線','none','JPR','環状八号線'
'none','none','000000',[],[consist,8],[consist,9],[consist,10],[consist,11],1).
frame(13,'交差','コウサ','JSA','交差'名詞-サ変接続,'none','3ccc14',[],[object,8],[object,12],1).
frame(14,'する','スル','JVB','する'動詞-自立,'基本形','000000',[],1).
frame(15,'交差する','none','JPR','交差する','none','none','000000',[],[consist,13],[consist,14],1).
...
frame(22,'本件事故現場','none','JNP','本件事故現場'
'none','none','000000',[],[consist,19],[consist,20],[consist,21],1).
frame(23,'交差点','コウサテン','JN1','交差点'名詞一般,'none','006646',[],[modifier,18],[a-object,22],1).
...
frame(29,'被害者','none','JNP','被害者','none','none','000000',[],[consist,27],[consist,28],1).
frame(30,'被告','ヒコク','JN1','被告'名詞一般,'none','10533b',[],2).
...
frame(33,'被害者太郎','none','JNP','被害者太郎','none','none','000000',[],[consist,31],[consist,32],2).
frame(34,'この','コノ','JNM','この'連体詞,'none','071b0d',[],2).
frame(35,'交点','コウテン','JN1','交点'名詞一般,'none','00D64c',[],[modifier,34],2).
frame(36,'運転','ウンテン','JSA','運転'名詞-サ変接続,'none','06c371',[],[agent,33],2).
frame(37,'する','スル','JVB','する'動詞-自立,'基本形','000000',[],2).
frame(38,'運転する','none','JPR','運転する','none','none','000000',[],[consist,36],[consist,37],2).
...
frame(41,'自動車','ジドウシャ','none','JNP','自動車','none','none','000000',[],[consist,39],[consist,40],2).
frame(42,'運転','ウンテン','JSA','運転'名詞-サ変接続,'none','06c371',[],[agent,30],2).
...
frame(45,'自動車','ジドウシャ','JN1','自動車'名詞一般,'none','0774c9',[],[modifier,44],2).
frame(46,'衝突','シヨウトツ','JSA','衝突'名詞-サ変接続,'none','3cc5d3',[],[agent,41],[object,45],[place,35],2).
...
frame(49,'負傷','フショウ','JSA','負傷'名詞-サ変接続,'none','106d85',[],[agent,29],[place,35],[sequence,48],1).
...
frame(52,'負傷した','none','JPR','負傷した','none','none','000000',[],[consist,49],[consist,50],[consist,51],2).

```

```

interRel述語
interRel(inter-equivalent, 1, 2).
interRel(inter-equivalent, 2, 3).
...

```

図 2 入力となる frame 述語と interRel 述語

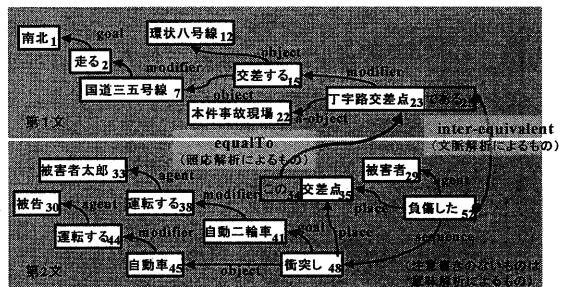


図 3 判例を表わす意味グラフ

2.4. 意味グラフ

先に述べた 4 つの自然言語処理を経て判例は、図 2 に示すように、文内の語の語意と係り受け関係にある 2 語間の文内深層格を表す frame 述語リストと、文間の文間深層格を表す interRel 述語リストに変換される。さらに、手作業で用意した照応解析データやラベル解析データ（詳細は 3 章で述べる）を併合して最終的に、図 3 に示すような意味グラフという形に変換される。ここでは、frame 述語が表すひとつひとつの語が頂点（ノード）で表現され、また、frame 述語間の文内深層格および文間深層格、照応解析・ラベル解析による深層格が弧（アーク）で表わされている。

意味グラフ中の各弧が表す深層格には、EDR 電子化辞書⁶にて定義された全部で 28 個の文内深層格に加え、「日本語文間の意味的關係解析システム InSeRA の開発研究」⁵における 21 個の文間深層格及び「日本語文書の意味解析シ

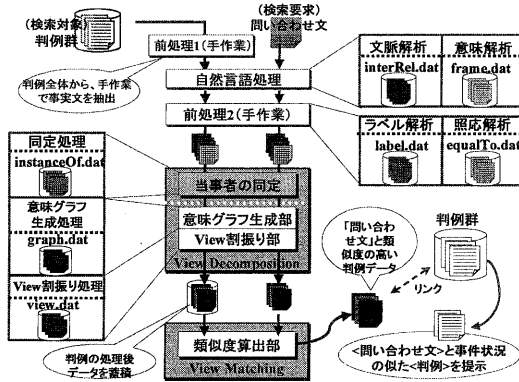


図 4 JCare の処理概要

テム SAGE の高速化と精度評価」⁴にて追加定義された 7 個の文間深層格が含まれる。

3. JCare の概要

システム JCare の処理の概要を図 4 に示す。まず前処理として「判例」から検索対象となる「争いのない事実文」を手作業で抽出し、自然言語処理にかける。「問い合わせ文」はそのまま自然言語処理にかける。この結果得られた意味グラフを View グラフに分割し、問い合わせ文と判例を View グラフレベルで照合し、類似度の高い順に判例を提示する。

3.1. 前処理

先に述べたように、照応解析とラベル解析は現状では手作業で行う。たとえば照応解析では、代名詞が文脈中に登場した人物を指す場合、図 5 のような equalTo 述語を出力する。

qualTo(照応元フレームID, [照応先フレームID, ...]).
equalTo(34, [23]).
equalTo(56, [53]).
equalTo(68, [59]).

図 5 照応解析結果を表わす equalTo 述語

第一 事案の概要
... 医療
本件は、幹線道路（環状八号線）の交差点において、仮免許を取得した者が運転し、右折しようとした数台目の普通乗用車と、その対向車線を走行してきた自動二輪車とが衝突し、自動二輪車の運転者が死亡した。
二 争いのない事実
1 当事者の地位
(1) 原告竹島邦治、同竹島品子、後記交通事故で死亡した竹島裕の父母である。原告竹島邦子は被害者の妻であり、同竹島裕子は被害者の子である。
(2) 被告芳野恵美は、仮免許を有し、被告株式会社ラヴィ...
... 後述省略...

図 6 ラベル解析結果を表わす label 述語

equalTo 述語は第 1 引数に照応元のフレーム ID を持ち、第 2 引数に照応先のフレーム ID 群のリストを持つ。

ラベル解析では、図 6 に示すような判例文における箇条書きの記述パターンを解析し、label 述語を出力する。label 述語は第 1 引数に文番号、第 2 引数に接頭語、第 3 引数に箇条書き番号、第 4 引数に接尾語、第 5 引数に箇条書き項目名をもつ。この第 2 から第 5 引数の記述パターンの解析から得た箇条書き構造により、どの内容文がどの箇条書き項目を説明したものなのかという情報を抽出する。

3.2. 当事者の同定

言語処理の出力データ群から意味グラフを作成する前に、当事者の同定を行う。当事者とは先に述べた通り事故に関係した人物であり、本研究では被告(defendant)、原告(plaintiff)、被害者(victim)、遺族(bereaved)、その他の当事者(others)の 5 つを用意する。これは当事者の記述表現が固有名詞であったり、被害者と原告が同一人物であるときに生じる照合時の当事者不整合を排除することが目的である。具体的には、被害者、被害者太郎など、同一人物について多様な表現を含む文章があったとき、これらを同一視するために図 7 に示す instanceOf 述語を出力する。これは第 2 引数のリスト中の各フ

instanceOf(同定クラス名,[フレームID, ...]).	
instanceOf(defendant,[30]).	< 被告
instanceOf(plaintiff,[]).	< 原告
instanceOf(victim,[29,33]).	< 被害者
instanceOf(bereaved,[]).	< 遺族
instanceOf(others,[]).	< その他の当事者

図 7 当事者同定結果を表わす instanceOf 述語

フレームが第1引数の当事者クラスに属することを表す。

3.3. 意味グラフ生成部

言語処理で出力する意味解析データ、文脈解析データに加え、前処理で用意する照応解析データ、ラベル解析データ、以上4つの解析データをまとめて、図3に示した意味グラフを生成する。具体的には図8に示す graph 述語で表す。graph 述語には頂点の情報を保持する graphNode 述語と、弧の情報を保持する graphArc 述語がある。それぞれの述語句名のあとには、検索質問(Query)か検索対象(Object)かを判別する接尾語が続く。graphNode 述語では第1引数にフレームID、第2引数に概念IDを保持する。また、graphArc 述語では引数を、係り元フレームID、深層格、係り先フレームIDの順で保持する。

弧(アーク)の情報を保持	頂点(ノード)の情報を保持
graphArcQuery(フレームID, 深層格, フレームNo).	graphNodeQuery(フレームID, 概念ID).
graphArcObject([2,goal,1]). graphArcObject([4,modifier,2]). graphArcObject([7,consist,3]). graphArcObject([7,consist,4]). ... graphArcObject([49,agent,29]). graphArcObject([49,place,35]). graphArcObject([49,sequence,48]). ...	graphNodeObject(2,'103ac6'). graphNodeObject(3,'3c3d32'). graphNodeObject(4,'0f38e8'). ... graphNodeObject(46,'3ce5db'). graphNodeObject(47,'000000'). graphNodeObject(48,'000000'). graphNodeObject(49,'106d85'). ...

図8 意味グラフの述語表現

4. View への分割

本研究では、意味グラフ間の類似度算出のために個々の語意だけでなく、語と語を結ぶ深層格も考慮する。一般に判例の事実文の意味グラフのサイズは頂点数が約1000個、弧数は約1220個であり、頂点と弧の両者を考慮して意味グラフ間の同型部分グラフ探索を行うと、計算量が

膨大になる。そのため探索空間を狭める方法として、意味グラフを視点 (View) により部分グラフ (View グラフと呼ぶ) 群に分割することにした。

4.1. View とは

View は、「事故現場の状況」「当事者の行動」「当事者の静的特徴」など、事実文中に記述されている概念の集まりを特徴付ける“視点”である。

本研究では、この View として“誰(who)”、“いつ(when)”、“どこ(where)”、“なに(what)”、“どのように(how)”という事件の基本構造 (Document Ontology) を表す Primary View と、「述語句」の意味により静的なもの動的なものへさらに細分割する Secondary View の2つを用意した。これらを判例用語で表わすと、表1に示す計11種の View となる。

1) Primary View

検索対象は、判例に含まれる「事件の状況を記述した事実文」であり、基本構造上の“when”、“where”、“what”では、これらの視点から選ばれたこの「事件」に直接に関連するものを対象とする。本研究では、対象が交通事故であることから、各々“when”は『発生日時』View、“where”は『発生場所』View、“what”は『関係車両』View と表現する。また“how”にあたる『事故態様』View は事件の状況をまとめて記述した (要約した) グラフをさし、たとえば「事故は被害者と被告が車を競わせたことでおきた。」などがこれにあたる。

“who”では『被告』View、『被害者』View、『原告』View の3つを用意する。先に述べた当事者の1つである「遺族」は、判例において存在しないことが多いと判断し、View からは外すことにした。また、事実文中に“被告”が複数人存在したとしても、被告らが関係する事実文は、同じ『被告』View へ割り振り、被告同士の区別は、意味グラフのレベルで行うこととした。以

表1 Primary View と Secondary View

Document Ontology	Primary View	Secondary View
when	発生日時	
where	発生場所	
who	加害者/被告	単独行動 静的特徴
	被害者	単独行動 静的特徴
	原告 (本人・遺族)	単独行動 静的特徴
what	関係車両	
how	事故態様	
others	その他	

上で述べた 6 つの View どれにも当てはまらないものは、『その他』View へ割り振る。

2) Secondary View

当事者に関する Primary View グラフを、「述語句」の意味によって、さらに Secondary View グラフへ分割する。一般的に「述語句」は、当事者の属性を示すものと、当事者の行動を示すものに大きく 2 種類に分けることが出来る。Secondary View は、この 2 種類の視点に異なった類似度算出方法を選択することで、視点の重みに応じた比較を行うために用意した。具体的には当事者に関する 3 つの View を、当事者の静的特徴についての記述を格納する『静的特徴』View と、当事者の行動についての記述を格納する『単独行動』View に分割する。たとえば、「被告は 20 歳の男性である」は前者に割り振られ、「被告は自動車を運転していた」は後者に割り振られる。また、行動に関する部分グラフの中に当事者が 2 人以上含まれていた場合は、『単独行動』View へ割り振らず、Primary View の『事故態様』View へ割り振る。これは、当事者別に View を分けたことで生じうる比較精度の低下を防ぐためである。たとえば、「問い合わせ文」での事故の状況を示す記述が、被告中心の表現となっており、反対に「判例」の「争いのない事実文」での記述が原告を中心の表現である場合、それぞれが異なる Secondary View へと割り振られてしまう。そのため、この事故の記述が似ているものであろうとなかろうと類似度はゼロとなる。このような照合の漏れを防ぐために、当事者が相互に関連しあいながら生じた行動を記述した部分グラフは、『事故態様』View へ割り振ることとする。

4.2. 主体と代表語

意味グラフ中の用言を表す頂点に接続している頂点は、この用言が表している対象でありこれを主体と呼ぶ。具体的には、EDR 電子化辞書の深層格で表される主体には、「有意志動作

を引き起こす主体」、「属性をもつ対象」および「動作・変化の影響を受ける対象」の 3 つがある。1 つ目は思考的・知的動作を含む意志を持って行われる動作の主体を表し、agent 格にあたるものである。2 つ目は、感情を表す動作の主体を表し、a-object 格にあたるものである。3 つ目は、自然現象・生理現象・物理現象の主体を表し、object 格にあたるのである。

さて実際に意味グラフの各頂点を View グラフに分配するには各頂点がどの View に属するかを判定する必要がある。この判定は、各々の View を代表する語(代表語)を用意し、意味グラフの各頂点が表す語がどの View の代表語の EDR 概念体系辞書中の下位概念であるかに基づいて行う。

表 2 Primary View の代表語

View	発生日時	発生場所	被告人	被害者	原告	関係車両	その他
代表語	時	場所	刑事被告人	被害者	原告	乗り物	左記以外
概念ID	<30f776>	<3aa938>	<0ef176>	<0eb47e>	<0efeb3>	<0f5354>	
代表語2			被告人	犠牲者	原告		
概念ID			<10533d>	<3cf862>	<3bd4e1>		
代表語3			被告	被害者			
概念ID			<10533b>	<3d023d>			

意味グラフ中の主体を表す各頂点を、Primary View に分類するための代表語を表にしたのが、表 2 である。これに対し Secondary View の代表語には図 9 に示すように、EDR 概念体系辞書の最上位概念群を利用し、『静的特徴』View では図中 A で表した計 5 つの概念を代表語とし、『単独行動』View では図中 B で表した計 4 つの概念を代表語として用意する。

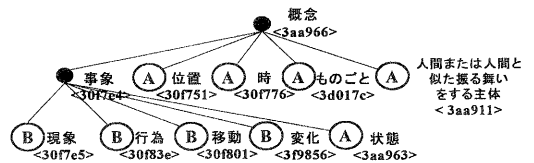


図 9 Secondary View の代表語

4.3. View への分割手法

意味グラフ生成部で生成した意味グラフを View に分配する際の単位は、「述語句」を 1 つ、「主体」を少なくとも 1 つ含む部分グラフとする。この単位で、意味グラフから述語句頂点に

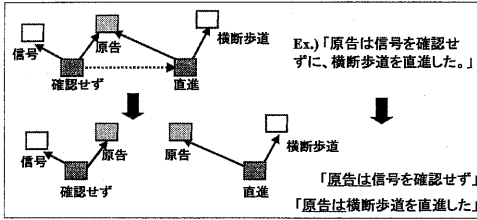


図 10 意味グラフの View への割振り単位

連結する頂点群を取り出し、先に述べたどの代表語の下位概念になっているかを判定し、適切な View へ分配する。意味グラフの切断箇所は「述語句」頂点間の弧とする。この際複数のグラフに共有されている頂点は、個々の View グラフにおいて重複して存在させることとする。たとえば「原告は信号を確認せずに、横断歩道を直進した。」という文章は図 10 のように、2つの View グラフとして扱われる。

5. View ごとのマッチング手法

問い合わせ文章と判例文章の各々 11 個の View グラフ同士を View ごとに両者間の位相同型部分の大きさをもとに、内容類似度を算出する。この手順について以下に説明する。

5.1. 語意類似度算出部

まず 2 つの View グラフから任意の頂点を取り出し、それらの語意類似度を算出する。語意類似度の値が一定値以上ならば仮のペアをつくる。すべてのノードどうしの判定終了後、ノードが重複した仮ペア群の類似度の最も高い

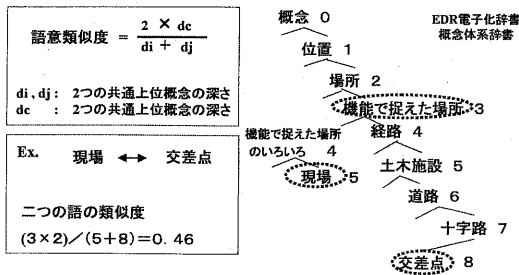


図 11 EDR 概念体系辞書を用いた語意類似度算出

ものを正式ペアとする。

2 ノード間の語意類似度の算出は、EDR 概念体系辞書を用いて 2 つの語が表す概念の概念体系木における深さと、2 つの共通上位語の深さを求め、図 11 左に示す式で算出する。たとえば“交差点”と“現場”というノード間の語意類似度は、図左下に示すように算出される。

正式ペアの語意類似度の合計を問い合わせ文のノード数で除算し、View グラフ各々の語意類似度を算出する。これらの合計値を総語意類似度とする。

5.2. 格類似度算出部

正式ペアになった頂点对から、任意の 2 つの頂点对を取り出し、その間にある最短路（関係パスと呼ぶ）上の深層格リストの類似性を評価する。この評価値を格類似度と呼ぶ。

具体的には、関係パス上の深層格群の類似性評価は、図 12 に示すようにグラフ 1 とグラフ 2 の概念対 A, B 間及び A', B' 間に有向路があれば、その最短路を求めその 2 つの関係パスが類似する深層格を共有することで得点を加算することで実現する。加える得点は深層格ごとに表 3 左に示すように定めた。この表で例えば、(agent, agent, 10)とは、「グラフ 1 とグラフ 2 の両方の関係パス上に深層格 agent が存在した時に、評価値 10 を加える」ということである。

概念対ごとにその関係パスに対するこの評価値の合計値を合計し、その結果を問い合わせ文の View グラフにおける格の数で除算することで、View グラフ各々の格類似度を算出する。さらに各 View グラフの格類似度を合計し総格

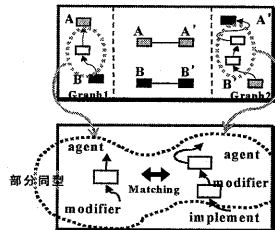


図 12 格類似度算出

表 3 深層格類似度評価値表(一部)

A	B
(agent, agent, 10)	(agent, agent, 10)
(object, object, 9)	(object, object, 9)
(cause, cause, 10)	(cause, cause, 10)
(agent, a-object, 8)	(agent, a-object, 8)
(a-object, agent, 8)	(a-object, agent, 8)
(reason, reason, 10)	(reason, reason, 10)
(sequence, sequence, 1)	(sequence, sequence, 8)
(when, when, 1)	(when, when, 8)
(during, when, 1)	(during, when, 7)
(during, during, 1)	(during, during, 8)
(inter-timingCondition, inter-timingCondition, 1)	(inter-timingCondition, inter-timingCondition, 8)
(inter-sequence, inter-sequence, 1)	(inter-sequence, inter-sequence, 8)

類似度を算出する。

ここで、関係パスにおける格類似性評価値を View ごとに変化させることで、全体の類似度を算出する上での View の独自性を考慮した。具体的には表 3 右に示すように、もう 1 つの評価値パターン B を用意した。ここではその一部を示す。これは主として時間軸における前後関係に関する格の評価値により多くの重みを与えたものである。パターン B は当事者の行動を記述した『単独行動』View グラフおよび『事故態様』View グラフ用に、パターン A はそれ以外の View グラフ用として利用する。

5.3. 意味グラフ類似度算出部

先で求めた総語意類似度と総格類似度を 11 種の View 毎に先の方法で合計した値を、全体の意味グラフ類似度とする。

2 つの処理を蓄積されている全ての判例に対して行い、判例各々の問い合わせ文に対する意味グラフ類似度を算出する。この類似度で判例をソートして、高いものからユーザへ提示することで、判例検索システムを実現する。

6. まとめ

本研究にて、文章を検索質問とした問い合わせシステムを提案し、具体的に判例検索システム JCare を開発した。また、意味グラフの View グラフへの分割と View グラフごとのマッチングによる類似度算出手法を提案し、検索の高速化・精度向上を図った。

実際に交通事故の 3 つの判例に対し問い合わせ文との照合を行ったところ、意味グラフ類似度 255.9, 163.42, 42.5 を得た。専門家に聞いた所、おおよそこの類似度は内容的な類似性の程度をよく表わしているとの評価を得た。

今後は類似度算出における関係パスの深層格の評価値を経験的に求めていく必要がある。また、前処理部などにおける手作業を自動化する必要がある。

謝辞

本研究を進めるにあたり、日本語形態素解析システム『茶筌』、係り受け解析システム『茶掛』を提供して下さった奈良先端科学技術大学院大学の松本裕治教授に深く感謝致します。また、判例検索システム構築において多くのアドバイスおよび判例データを提供して下さった第一法規株式会社に深く感謝致します。

【参考文献】

- [1]日笠亘,藤井綱貴,黒橋禎夫:”入力質問と知識表現の柔軟なマッチングによる対話的ヘルプシステムの構築”, 自然言語処理研究報告 No134, pp. 134-14(1999.11).
- [2]松本祐治,北内啓,山下達雄,平野善隆,今一修,今村友明: 日本語形態素解析システム『茶筌』version2.0 使用説明書,奈良先端科学技術大学院大学松本研究室(1999).
- [3]原田実,水野 高宏: “EDR を用いた日本語意味解析システム SAGE”, 人工知能学会論文誌, Vol.16, No.1, pp.85-93 (2001.1).
- [4]田淵和幸,原田実:” 日本語文書の意味解析システム SAGE の高速化と精度評価”, 言語処理学会 第 7 回 年次大会 論文集 (印刷中), (2001. 3).
- [5]川端崇央,原田実:” 日本語文間の意味的關係解析システム InSeRA の開発研究”, 情報処理学会, 自然言語処理研究会論文集 (印刷中), (2001. 3).
- [6]㈱日本語電子化辞書研究所: EDR 電子化辞書使用説明書, ㈱日本語電子化辞書研究所 (1995).