

述語項構造を用いた単純な司法試験自動解答器の構築

Naive Legal Yes/No Question Answering System using Predicate Argument Structures

谷口 諒輔[†]
Ryosuke Taniguchi

狩野 芳伸[†]
Yoshinobu Kano

1. はじめに

法律に関するテキストの質問応答は、他の領域の質問応答とは大きく異なる。他領域における質問応答システムとの最大の差異は、法律文書に登場する複数のエージェントの役割とその関係性を、正確に解析する必要があるということである。我々のシステムでは、与えられた司法試験問題及び知識源となる民法条文の間の役割と関係性を対応づけるために、述語項構造をベースとして、法律関係文書に登場する格の役割を解析し、それを利用する。我々は本システムを、国際コンペティションである Competition on Legal Information Extraction/Entailment 2016 (以下 COLIEE)[1]に適用し、サブタスクの Phase 2 にて、過去のデータセットの解析結果ではどのチームよりも良い結果を、また formal run では 1 位タイの結果を得た[2]。これらの結果より、法律文書において、上記に示した格の役割とその関係性の重要性が示されたが、今後、より高い正答率・確信度を得るために、幾つかの問題が残されている。本稿では、COLIEE に参加するにあたり用いた手法、及びその追加的な解析と、残された課題についての改善について述べる。

t1: (留置権の行使と債権の消滅時効)
第三百条 留置権の行使は、債権の消滅時効の進行を妨げない。
(Exercise of Rights of Retention and Extinctive Prescription of Claims) Article 300
The exercise of a right of retention shall not preclude the running of extinctive prescription of claims.
t2: 留置権者が留置物の占有を継続している間であっても、その被担保債権についての消滅時効は進行する。
Even while the holder of a right to retention continues the possession of the retained property, extinctive prescription runs for its secured claim.

図 1. COLIEE データセット例

2. COLIEE タスク

COLIEE シリーズは、JURISIN (Juris-Informatics) workshop において催されており、本稿が対象とするのは COLIEE 2014[3]、COLIEE 2015[4]に続く 3 度目の COLIEE 2016[1]である。

COLIEE のタスクには 3 つの Phase (サブタスク) が設定されている。Phase 1 では、与えられた司法試験問題を解答するのに必要な民法条文を検索することを目的としている。

Phase 2 では、含意関係認識のタスクを担う。図 1 は COLIEE におけるデータセットの一例を示す。試験問題 (t2 に対応する) とそれに関する民法条文 (t1 に対応する) が与えられ、システムは Yes か No の二値で与えられた問題文が、民法条文に照らして正しいか否かを解答する。

[†] 静岡大学大学院総合科学技術研究科情報学専攻

Phase 3 では試験問題 (t2) のみが与えられ、システムはそれに関する民法条文を全条文から検索したのち、含意関係を判断しなければならない。

問題文 (t2) は、我が国の司法試験のうち短答式問題の民法分野のものが用いられ、また知識源として与えられる民法条文も我が国の民法通則である。COLIEE ではそれぞれを英語に翻訳したのもも提供されたが、我々は日本語版を対象としてシステムを構築する。また我々は主として Phase 2 を対象としたシステムを設計し、それを転用する形で Phase 3 の解答を試みる。

3. 提案手法

3.1. 設計思想

我々は、ルールベースで構築されたシンプルな解答器を目指す。その理由として、与えられる民法条文及び司法試験問題の各文構造が形式的で、主題、条件節、対象などが、ある程度決められた格や表現を伴い、文中に登場していることが挙げられる。また、法律文書の観察から得られた知見を元に、ルールベースでシステムを作成することによって、法律文書が内包する問題を明らかにするためである。

法律文書においては、格の役割とその関係が文構造を決定する上で重要であると考え、本稿では表層格を用いて主題を明らかにし、それと結びつく述語、とりわけ文末表現の関係性に着目する。

解析には、日本語格の依存関係を明らかにするために、日本語形態素解析器及び日本語構文解析器を使用する。これらの解析器で得られた結果を利用することで、主題と述語を、それぞれの文で得ることが可能となる。現状、より意味的な解釈が可能な深層格を得られる解析器は存在しないため、ここでは表層格を用いることとする。主語と認定する際に必要な主格は、各文に一般的に登場するガ格とハ格であると定義した。本システムは主語と主題についての明確な区別は行わないものとする。

条文の冒頭部分には、その条文の条文番号、及び条文明が記載されている。民法条文を解析する際、含意関係認識に不要だと考えられる、各文の冒頭部分に記載される「X 条」とそれに付随する文章は事前に取り除く。さらに、補足・逆説を表現する接続詞である「ただし」が文中に出現した場合は、これを含む一文を取り除いて処理する場合、そのままの状態でも処理する二つの方法を用意した。「ただし」は付加的な接続詞であるものの、文全体の論理値を逆転させる可能性もあるためである。

以下の節で、主題と文末表現を 1 対 1 で対応させ比較するメソッド 1、及び主題と複数の文末表現を 1 対多で対応させ比較するメソッド 2 について説明する。

3.2. 主題及び文末表現の 1 対 1 対応付け (メソッド 1)

メソッド 1 では、前述した手法により抽出した主題と文末表現の関係づけを行う。図 2 にその概念図を示す。

第一に、構文解析及び形態素解析の結果から得られた主題と文末表現について、一対一の対応付けを行う。今回、文中に主題となり得る表現が存在しなかった場合、またはそれを取得できなかった場合は、文末表現のみを獲得し、それらを比較する。

第二に、得られた主題と文末表現のペアについて、より抽象的な形へと変形することを試みる。例えば、主題となり得る単語に付随する表層格(ガ格/ハ格)は、ここで取り除く。同時に、句読点についてもここで取り除く。さらに、文末表現が可能性表現(～できる/～できない)であった場合、その文末表現部分を True、もしくは False へと置換する。文末表現に関して、可能性表現ではないとき、文末表現は表層形のままを維持する。メソッド 1 では民法条文及び試験問題から獲得された全ての主題及び文末表現のペアが一致するものがあれば Yes と解答し、そうでなければ No と解答する。

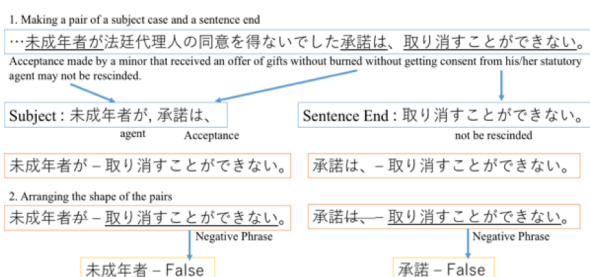


図 2. メソッド 1 の概念説明

3.3. 主題及び文末表現の 1 対多の対応付け (メソッド 2)

メソッド 2 はメソッド 1 と比較し、より厳格な評価を期待した手法であり、一つの主題に対して複数の文末表現を持たせることが可能となる。メソッド 2 ではまず、ペアに含まれる主題を比較し、もし一致する表現が見つければ、そのセット中の文末表現部分の比較を行う。図 3 は、メソッド 2 の概念を示したものである。例えば、「善意の占有者は、占有物から生ずる果実を取得する。善意の占有者が本権の訴えにおいて敗訴したときは、その訴えの提起の時から悪意の占有者とみなす。」という文において、我々のシステムは「占有者は」を主題、「取得する」及び「みなす」を述語として獲得することができる。

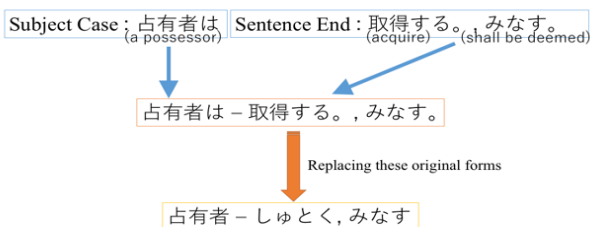


図 3. メソッド 2 の概念図

4. 実験

解析には、日本語形態素解析器 JUMAN[5]と、日本語構文解析器 KNP[6]を使用する。また JUMAN に、本タスクに向けて日本語法律用語辞典「有斐閣法律用語辞典第 4 版」をユーザ辞書として追加する。法律用語辞典で追加される

単語数は延べ 12,612 単語である。KNP は日本語格の依存関係を解析するツールとしても利用可能であり、この際に JUMAN の形態素解析結果を用いる。

実験は COLIEE2016 における日本語版データセットを用いて行った。データセットは訓練・テストの各問題データと民法通則条文からなり、訓練は過去の司法試験問題を COLIEE 形式にしたのち正答を付したものである。訓練データは、参考とすべき民法条文(t1)と試験問題(t2)が各問にセットに含まれており、その数は平成 21 年度が 56 問、平成 22 年度が 47 問、平成 23 年度が 41 問、平成 24 年度が 79 問、平成 25 年度が 65 問である。またテストデータは、Phase 2 用データは訓練データと同じデータフォーマットであり平成 27 年度のもので 70 問、Phase 3 用データは情報検索タスクである Phase 1 も含むため民法条文(t1)は記載されておらず、平成 26 年度の 95 問となっている。COLIEE の Phase 2 および Phase 3 における評価は accuracy が用いられている。

我々の手法では、システム全体のパフォーマンスのチェック、及び機能向上のための解析にこれらの訓練データを利用している。

4.1. 訓練データにおける Phase 2 の結果

Phase 2 では、2 つのメソッド及び接続詞「ただし」の扱い方により、4 つの Rule (実行パターン) を作り実行した。Rule1 ではメソッド 1 とメソッド 2 を組み合わせて利用している。Rule1 では全ての文章を対象に解析を行った。Rule2 は Rule1 とほぼ同様だが、接続詞「ただし」を含む文については除外し解析を行う。Rule3 及び Rule4 ではメソッド 1 のみを適用し、「ただし」を含む一文も処理するものを Rule3、除外するものを Rule4 とした。

Dataset	H21	H22	H23	H24	H25
Rule1_Accuracy	66.07	63.83	65.85	67.09	72.73
Rule1_F-measure	66.06	63.76	65.67	66.44	72.32
Rule2_Accuracy	60.71	61.70	65.85	67.09	74.24
Rule2_F-measure	60.71	61.55	65.67	66.44	74.09
Rule3_Accuracy	64.29	68.09	60.98	73.42	71.21
Rule3_F-measure	64.10	68.03	60.95	73.55	70.67
Rule4_Accuracy	60.71	65.96	60.98	70.89	72.73
Rule4_F-measure	60.66	65.94	60.95	70.72	72.50

表 1. 法律用語辞典を用いたフェーズ 2 結果
色つきのセルは該当年度のベストスコアを示す

表 1 に、Phase 2 における訓練データを利用した際の評価結果を示す。前年度の COLIEE 2015 におけるテストデータであった平成 25 年度のデータセットについて、COLIEE 2015 の参加者のベストスコアが 66.67%であったのに対して、我々のシステムは 74.24%であり、約 7 ポイント上回ることができた。

4.2. 訓練データにおける Phase 3 の結果

Phase 3 では、Phase 2 と同様のシステムを使用した。Phase 2 は、与えられた t2(民法条文)を知識源とし解答する含意認識タスクである。一方 Phase 3 は、t2 は与えられない。このため、我々のシステムは民法全てを解析の対象とした。このとき、それぞれの試験問題に対し幾つの条文が該当するか、言い換えるに Yes をいくつ返すかをカウントする。そして Yes の数を閾値と比較し、最終的な Yes もしくは No の解答を決定している。表 2 は訓練データを用い

た際の結果を示している。ただし、Phase 3 においては「ただし」を含む文章の有無はスコアに全く差がなかったため、表中では Rule 2 および Rule 4 の結果は記載せず、Rule 1 及び Rule 3 の結果のみ記載した。閾値については、実験的に決定した 50 に加え、訓練データを解答する際は、解答する年度を除いた他 4 年度分のデータから得られる Yes の数を母集団とし計算された、中央値と平均値を用いる。テストデータを解く際には、訓練データ全体から得られる中央値と平均値を用いた。

Threshold	Dataset	H21	H22	H23	H24	H25
50	Rule1_Accuracy	48.21	48.94	56.10	58.23	43.94
	Rule1_F-measure	48.20	45.98	56.07	58.20	43.93
	Rule3_Accuracy	48.21	46.81	58.54	46.84	43.94
	Rule3_F-measure	39.74	41.63	53.06	39.80	36.19
Median	Rule1_Accuracy	48.21	48.94	58.54	54.43	43.94
	Rule1_F-measure	48.20	45.98	58.54	54.36	43.93
	Rule3_Accuracy	53.57	42.55	51.22	58.23	36.36
	Rule3_F-measure	53.51	41.60	51.19	58.20	36.31
Average	Rule1_Accuracy	46.43	51.06	53.66	55.70	42.42
	Rule1_F-measure	46.15	44.65	53.21	55.44	41.94
	Rule3_Accuracy	57.14	53.19	51.22	54.43	40.91
	Rule3_F-measure	56.65	50.48	51.19	54.42	40.90

表 2. 訓練データの Phase 3 実行結果
色つきのセルは各年度のベストスコアを示す。
"Median" は中央値、"Average" は平均値である。

4.3. COLIEE2016 における formal run の結果

テストデータによる評価である formal run も、訓練データ解析時に用いた Phase 2 及び Phase 3 のシステムと同じものを使用した。表 3 は Phase 2 の実行結果を、表 4 は Phase 3 における実行結果を示している。それぞれの表における KIS-1 は Rule1、KIS-2 は Rule2、KIS-3 は Rule3、KIS-4 は Rule4 を適用したものである。他の Run は COLIEE 2016 に参加した他チームの結果である。

KIS-1 及び KIS-2 はメソッド 1 とメソッド 2 を利用したものであり、Phase 2 ではいずれの他チームの結果よりも上回っていることわかる。また Phase 3 における KIS-3 及び KIS-4 はメソッド 1 のみを適用したもののだが、KIS-1 及び KIS-2 を上回っていることがわかる。

我々の手法は Phase 2 において、深層学習を用いた iLis7[10]と並び 1 位タイとなり、Phase 3 では 3 位となった。

5. 考察

第一に、我々は形態素解析器を用いる際、法律用語辞典をユーザ辞書として追加した。図 4 は、法律用語辞典の適用の有無より得られた、処理の差異を示したものである。この図において、法律用語辞典を用いた Rule3 は Rule3A、法律用語辞典を用いず、JUMAN 標準の辞書のみを利用した Rule3 は Rule3B と示す。図 4 の例 H22-9-I について、形態素解析において法律用語辞典を用いない Rule3B は、「付合」という単語が、「付」及び「合」に分解されることがわかる。また「付」は、文末表現と結びつき、その結果間違った Yes を出力していることがわかる。図 4 の別の例 H22-23-O は、その逆のケースと言える。法律用語辞典なしで解析した結果、「貸借人」は「貸借」と「人」に分解される。これは「人」がより一般的な単語であることに起因すると予想され、またそれは他の文末要素と紐付けされた。この結果、システムは誤った回答である No を出力した。これらの結果を省みるに、法律用語辞典は効果

Run	Accuracy	Run	Accuracy
JNLN1 [7]	52.86	UofA [10]	55.71
KIS-1	62.86	N01-1 [11]	54.29
KIS-2	62.86	N01-2 [11]	54.29
KIS-3	58.57	N01-3 [11]	48.57
KIS-4	58.57	N01-4 [11]	48.57
iLis7 [9]	62.86	N01-5 [11]	57.14
JNLN3 [8]	48.57		

表 3. formal run の Phase2 実行結果
色つきのセルは COLIEE 2016 参加者中のベストスコアを示す

Run	Accuracy	Run	Accuracy
JNLN1 [7]	40.00	iLis7 [9]	53.68
KIS-1	51.58	JNLN3 [8]	47.37
KIS-2	51.58	UofA [10]	46.32
KIS-3	52.63	UofA [10]	54.74
KIS-4	52.63	UofA [10]	55.79

表 4. formal run の Phase3 実行結果
色つきのセルは COLIEE 2016 参加者中のベストスコアを示す

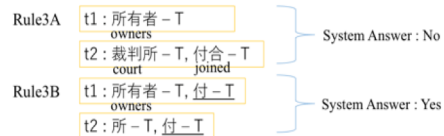
Threshold	Dataset	formal run	Threshold	Dataset	formal run
Median	Rule1_Accuracy	51.58	Average	Rule1_Accuracy	50.53
	Rule1_F-measure	42.92		Rule1_F-measure	38.51
	Rule3_Accuracy	48.42		Rule3_Accuracy	48.42
	Rule3_F-measure	32.62		Rule3_F-measure	32.62

表 5. 中央値及び平均値を用いた際のテストデータ実行結果

的な、言い換えるに、我々が意図する主語と述語の対応を抜き出す補助的な役割を果たしたと言える。

第二に、Phase 2 において、平成 25 年度の訓練データでは、「ただし」を含む文を除外して解析した Rule1 及び Rule3 が、Rule2 及び Rule4 よりも良い結果を示したことが読み取れる。「ただし」という接続詞を含む一文は、補足及び逆説の意味的な立場から、元来とは真逆の結果に導く可能性がある。このことから、それを取り除いたことによって、よりよい結果を得られたと考えられる。

H22-9-I (Correct Answer : No)



H22-23-O (Correct Answer : Yes)

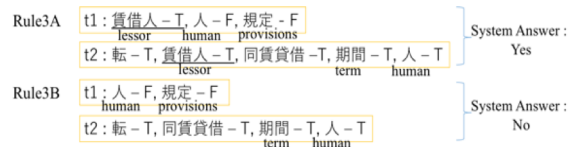


図 4 法律用語辞典の効果
Rule3A は法律用語時点を用いたもの、Rule3B は用いないものを示す

第三に、我々のシステムは Phase 3 を解答する際に、Phase 2 のシステムを直接用いたにも関わらず、表 3 及び表 4 より、Phase 2 における結果の良し悪しは、Phase 3 に相関しないということが読み取れる。これは、システムが Phase 3 ではなく、Phase 2 に最適化されていることが一因であると考えられる。また表 2 より、Rule3 について、50 を閾値としたときと比べ、中央値または平均値を閾値として用いたときは、精度と F 値の値に差がでないことがわかる。Rule1 と Rule2 のそれぞれに設定された中央値や平均値によって、Yes/No の出力バランスが整えられたと考えられる。対して、表 5 は、中央値及び平均値を用いた Phase 3 の解答システムをテストデータに適用した結果であり、表

2 と比べて精度と F 値に大きな差が見られる。これにより、Phase 3 の formal run に用いられたテストデータと訓練データに傾向差があったことが示されている。

また、我々がシステムのパフォーマンス向上のために使用した平成 21 年度から平成 25 年度までの訓練データについて、評価結果には年度によるばらつきがあることが見て取れる。Phase 2 タスクの全体的な傾向として、表 3 の結果が表 1 よりも下回っていることがわかるが、システムにとってこのときのテストデータが、訓練データよりも解き難いものであったことは明らかである。このことから、配布されたデータセットについて、年度ごとに傾向差があることが予想される。このようなスコアのばらつきが如何にして引き起こされたか、各データについてより正確に解析を進めていく必要がある。

さらに、我々のシステムの精度は、日本語形態素解析器である KNP に依存しているという点についても言及したい。KNP は、大規模ウェブテキストを用いて学習されており、法律文書における精度は、確かなものであるとは言えない。

Phase 2 において、同率 1 位であった iLis7 は、決定木を用いた多数決法、線形 SVM、語意、意味、また構文といった様々な特徴量を用いた、畳み込みニューラルネットワークを使用している。しかし実際には、機械学習の効果は Yes/No のバランスの学習と、語彙レベルの類似を利用していると考えられ、法律文書の複雑な問題を解くために必要不可欠な特徴を捉えられているとは考え難い。COLIEE 2016 にて配布されたデータセットは、教師付き機械学習を用いた手法を採用するには不十分な数である。同時に、end-to-end の教師付き機械学習手法の適用だけでは、民法条文及び試験問題における文構造や照応関係を明らかにするための、より深層的な特徴は取得することは困難であると考える。

6. 今後の課題

過去の COLIEE に参加したシステムと比較した際の最大の相違点は、我々が日本語版データセットを利用していることに対し、他のチームは英語版データセットを利用しているということである。英語版データセットは、元となる日本語の司法試験問題及び民法条文を直接翻訳したものであるが、日本語テキストを扱う際には、形態素に分割しなければならない。この形態素解析の際にシステムが誤れば、最終的な結果も予期したものと異なったものになり得る。我々が観察する限り、法律用語辞典を用いた形態素解析結果はある程度、信用できるものであるといえるが、幾らかの問題については、形態素解析の誤りも認められる。英語版データセットと日本語版データセットの比較が今後課題といえる。

過去に行われた COLIEE を省みるに、高い精度を十分に維持しつつ論理的関係性を抽出することは極めて難しいと考えられる。例えば、法律文書においては、「実体化」の概念を含む問題がある。本稿で述べる「実体化」とは、具体的な行為者が、アルファベットを用いて代名詞化されている問題を指す。「実体化」を伴う問題の分別は、より高度な抽象的処理過程を経なければならないだろう。

知識源である民法条文は、論理や条件が一つの文中に、もしくは断片的に散りつつ記されている。また、民法条文は独特の否定表現や省略表現が含まれており、これらは問題を解くにあたって極めて重要な要素である。我々のシス

テムではある程度、それらの表現について取り扱っているが、全てを網羅できてはいない。

評価結果は、我々のシステムの有効性をある程度示している一方で、JUMAN 及び KNP の解析結果から取得可能な、表層格のみを取り扱っていることに着目したい。より深層的な格役割を担う深層格を用いることは、本タスクで使用したような法律文書における意味の関係性を明らかにする上で、極めて重要であると予想される。

前述した問題点は、将来的な研究で解決することが望まれる。もっとも本質的かつ重要な問題点は、論理的関係性と抽象化問題を解決できるか否かである。

7. 結論

本稿では、我が国における司法試験、短答式問題の民法分野を解くことを目標とする COLIEE 2016 を対象に、我々が提案した手法と実装、実験結果について述べた。法律文書の文構造の観察に基づいて、我々のシステムは、司法試験問題と知識源に当たる民法条文における、格の役割とその関係性を明らかにするために、述語項構造を利用した。また、我々は本システムを、COLIEE 2016 で配布された日本語版データセットに適用した。我々のシステムは、前年度開催された COLIEE 2015 参加者のベストスコアを 7 ポイント上回り、また COLIEE 2016 の formal run では一位タイのパフォーマンスを示した。これは、我々が着眼した開発方針の重要性を示している。

参考文献

- [1] M.-Y. Kim, R. Goebel, Y. Kano, and S. Ken, "COLIEE-2016: Evaluation of the Competition on Legal Information Extraction/Entailment," in Tenth International Workshop on Juris-informatics (JURISIN 2016), 2016.
- [2] Ryosuke Taniguchi and Yoshinobu Kano, "Legal Yes/No Question Answering System using Case-Role Analysis," in Tenth International Workshop on Juris-informatics (JURISIN 2016), 2016.
- [3] "Competition on Legal Information Extraction/Entailment (COLIEE-14), Workshop on Juris-informatics (JURISIN) 2014," 2014. [Online]. Available: http://webdocs.cs.ualberta.ca/~miyoung2/jurisin_task/index.html.
- [4] M.-Y. Kim, R. Goebel, and S. Ken, "COLIEE-2015: Evaluation of Legal Question Answering," in Ninth International Workshop on Juris-informatics (JURISIN 2015), 2015.
- [5]
- [6] "JUMAN (a User-Extensible Morphological Analyzer for Japanese)." [Online]. Available: <http://nlp.ist.i.kyoto-u.ac.jp/EN/index.php?JUMAN>.
- [7] "Japanese Dependency and Case Structure Analyzer KNP." [Online]. Available: <http://nlp.ist.i.kyoto-u.ac.jp/EN/index.php?KNP>.
- [8] Danilo S. Carvalho, Vu Duc Tran, Khanh Van Tran, Viet Dac Lai, and Minh-Le Nguyen, "Lexical to Discourse-level Corpus Modeling for Legal Question Answering," in Tenth International Workshop on Juris-informatics (JURISIN 2016), 2016.
- [9] Phong-Khac Do, Huy-Tien Nguyen, Chien-Xuan Tran, Minh-Tien Nguyen and Nguyen Le Minh, "Legal Question Answering using Ranking SVM and Deep Convolutional Neural Network," in Tenth International Workshop on Juris-informatics (JURISIN 2016), 2016.
- [10] Kiyoun Kim, Seongwan Heo, Sungchul Jung, Kihyun Hong and Young-Yik Rhim, "An Ensemble Based Legal Information Retrieval and Entailment System" in Tenth International Workshop on Juris-informatics (JURISIN 2016), 2016.
- [11] Mi-Young Kim, Ying Xu, Yao Lu and Randy Goebel, "Legal Question Answering using Paraphrasing and Entailment Analysis," in Tenth International Workshop on Juris-informatics (JURISIN 2016), 2016.
- [12] Adebayo Kolawole John, Luigi Di Caro, Guido Boella, and Cesare Bartolini, "TEAM-NORMAS' PARTICIPATION AT THE COLIEE2016 BAR LEGAL EXAM COMPETITION," in Tenth International Workshop on Juris-informatics (JURISIN 2016), 2016.