機関番号	研究種目番号	応募区分番号	小区分	整理番号
13901	05	1	62020	0001

令和3(2021)年度　基盤研究（B）（一般）研究計画調書

令和 2年10月20日
1版

新規

研究種目	基盤研究(B)		応募区分	一般			
小区分	ウェブ情報学およびサービス情報学関連						
研究代表者 氏名	(フリガナ)	コマミズ タカヒロ					
	(漢字等)	駒水 孝裕					
所属研究機関	名古屋大学						
部　局	情報基盤センター						
職	助教						
研究課題名	異種オープンデータ活用のためのデータ統合・管理基盤の研究開発						
研究経費 （千円未満の端数は切り捨てる）	年度	研究経費（千円）	使用内訳（千円）				
			設備備品費	消耗品費	旅費	人件費･謝金	その他
	令和3年度	5,350	4,750	0	100	300	200
	令和4年度	4,050	300	50	2,400	500	800
	令和5年度	6,250	2,300	50	2,400	600	900
	令和6年度	4,350	300	50	2,400	600	1,000
	令和7年度	0	0	0	0	0	0
	総計	20,000	7,650	150	7,300	2,000	2,900
開示希望の有無	審査結果の開示を希望する						
研究計画最終年度前年度応募	--						

研究組織（研究代表者及び研究分担者）

	氏名（年齢）	所属研究機関 部局 職	学位 役割分担	令和3年度 研究経費 (千円)	エフォート (%)
研究代表者	30756367　　() コマミズ　タカヒロ 駒水　孝裕	名古屋大学 情報基盤センター 助教	博士（工学） 総括，データ統合	3,400	30
研究分担者	80314532　　() ハタノ　ケンジ 波多野　賢治	同志社大学 文化情報学部 教授	博士（工学） 統合データ検索	1,050	5
研究分担者	10332157　　() イデ　イチロウ 井手　一郎	名古屋大学 数理・データ科学教育研究センター 教授	博士（工学） マルチメディアコンテンツ処理	900	10
合計　　3 名			研究経費合計	5,350	

１　研究目的、研究方法など

本研究計画調書は「小区分」の審査区分で審査されます。記述に当たっては、「科学研究費助成事業における審査及び評価に関する規程」（公募要領１１１頁参照）を参考にすること。
本欄には、本研究の目的と方法などについて、４頁以内で記述すること。
冒頭にその概要を簡潔にまとめて記述し、本文には、(1)本研究の学術的背景、研究課題の核心をなす学術的「問い」、(2)本研究の目的および学術的独自性と創造性、(3)本研究で何をどのように、どこまで明らかにしようとするのか、について具体的かつ明確に記述すること。
本研究を研究分担者とともに行う場合は、研究代表者、研究分担者の具体的な役割を記述すること。

（概要）

【背景】オープンデータ化が進み，さまざまなデータが公開されてきた．データの種類もテキストからマルチメディアと多様になり，かつそれぞれが Web 上に散在している．そのため，異種データを横断的に利用するには，データを収集し，相互の関連性を構造化することが必要となる．
【目的】本研究では，LOD (Linked Open Data) を起点にマルチメディアを含む異種フォーマットのオープンデータ統合・管理に関する研究分野 Linked Open Multimedia Data Management を創造する．本研究はその一端として，データ統合とデータ横断的検索技術の確立を目的とする．
【展開】本研究では，以下の課題に取り組む．(1) **異種データに共通するエンティティの同定**：従来対象とされてこなかったマルチメディアへの対応とデータ間における情報の偏りによる性能低下への対処に取り組む．(2) **異種データ横断検索エンジンの構築**：本研究では，LOD における標準問合せ言語 SPARQL を，異種データを扱えるように問合せ処理エンジンを拡張する．

（本文）

(1)　本研究の学術的背景、研究課題の核心をなす学術的「問い」

【学術的背景】

インターネットの普及や各種デバイスの大衆化により，電子的なデータは急速に増加している．データの急増と計算機デバイスの高性能化・低廉化に後押しされ，人工知能技術を筆頭にデータ科学技術は大きく進歩してきた．この背景には，オープンサイエンスやオープンガバメントなどの根底にあるオープンデータの動きがある．データをオープン化することで，得られた知見や知識，主張の透明性や再現性を担保できるため，オープンデータは社会的に重要な動きである．さらに，データの相互利用性を高めるために，Linked Open Data (LOD) という枠組みが World Wide Web を提唱した Tim Berners-Lee 氏によって提唱されている．

LOD は研究データを中心に急速に普及しつつある[1]が，公共機関のオープンデータは独立に作られ公開されているため，その相互利用性は低い．その主な要因としては，LOD が要求するデータ形式 (RDF; Resource Description Framework) への変換が容易でないこと，他のデータと結びつける労力が大きいことの 2 つが挙げられる．そのため，オープンデータを相互利用するためにデータ同士を結びつける人的作業が必要になる．この現状を打開する一つの方法は，多様なフォーマットのオープンデータを相互利用できるように，データの統合を支援することである．

オープンデータのほとんどは構造化されたテキストであるが，画像や映像などのマルチメディアもまた重要な知識を含むデータである．例えば，人物や建物の画像は被写体の特徴を視覚的に表現し，自然現象や事件・事故などを観測した画像や映像は起こったことを視覚・聴覚的に記録した知識である．LOD はテキストで記述される知識を構造化するため，マルチメディアから抽出した（テキスト化された）知識は扱えるが，マルチメディア自体はそれを識別する文字列によって表現される．そのため，マルチメディアが内包する意味や知識をもれなく表現することはできない．これに対し，マルチメディアを知識として LOD に取り込み，マルチメディアとテキストを相互利用できる環境が実現されれば，オープンデータを活用したアプリケーションの作成やデータ工学技術の研究開発をより促進することが期待される．そのためには，オープンデータ管理においてテキストをもとに構造化されたデータに限定した方法論から脱却しなければならない．

[1] https://lod-cloud.net/

【1　研究目的、研究方法など（つづき）】

【学術的「問い」】

本研究における学術的な問いは「多種多様なオープンデータを統合的に管理し，効率的に活用するためのデータ管理基盤を構築できるか」である．上述のように，オープンデータは Web 上に散在していることに加え，種類が多様化してきた．本研究は，これらのデータを相互にリンクし，横断的な活用を容易にするデータ管理基盤を実現することで，異種データを活用した技術の研究開発および実用化の促進を目指すものである．

(2)　本研究の目的および学術的独自性と創造性

【本研究の目的：Linked Open Multimedia Data Management 領域の創造】

上記の問いに対し，本研究は多種多様なオープンデータを統合・管理し，横断的な異種データ活用およびデータや技術の循環（データや技術の更新や修正）を実現できるプラットフォームを構築する．これを実現するために，本研究では他のデータとの接続が容易な LOD を中心に据える．LOD を中心に，異種データを統合・管理し活用するための研究はこれまでに行われておらず，新たな研究領域を開拓する必要がある．本研究は，この新領域「**Linked Open Multimedia Data Management**」を創造するものである．

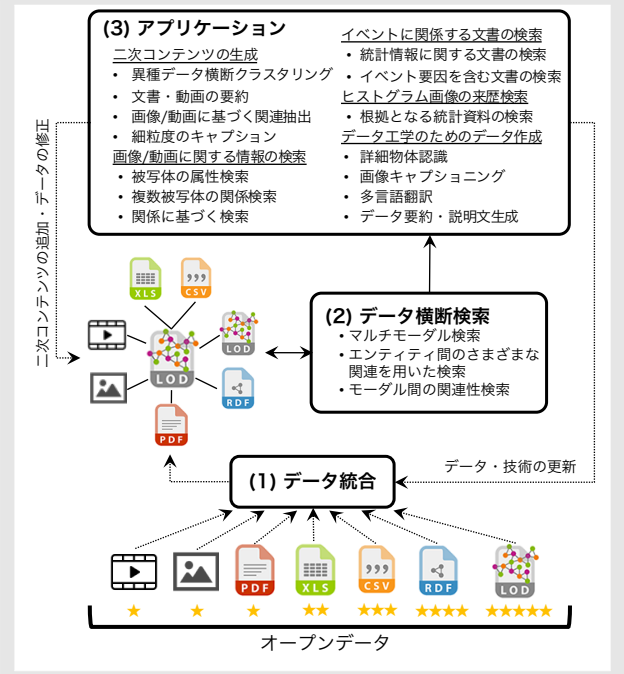

図 1: Linked Open Multimedia Data Management

この新領域を構成する要素を図 1 に示す．主な構成要素は，(1) データ統合，(2) データ横断検索，およびそれらの結果を用いた (3) アプリケーション開発，である．(1) では，他のドメインの LOD およびさまざまなフォーマット（テキスト，表形式，画像，映像，など）のデータを LOD のエンティティに紐付けることで，異種データを統合する．(2) では，統合されたデータから必要なデータを検索する．(3) では，統合されたデータを活用し，データ工学技術を利用してより付加価値の高いデータを生成したり，異種データ横断的なアプリケーションを開発したり，多様かつ大量のデータを供給することでより高度な技術開発を行う．(3) で生成した二次的なデータを統合データに加えることで，単にオープンデータを結合した以上の付加価値の高い統合データを構築する．また，(3) のアプリケーション開発の過程で発見されたデータの不具合を統合されたデータに常にフィードバックすることでより精緻なデータを作り上げる．さらに，(3) の結果は，(1) のデータ統合に対してデータの追加や新たな統合技術としてフィードバックされ，データ統合がより高精度なものとなる．

【学術的独自性と創造性】

本研究は，さまざまなフォーマットで公開されるデータを統合的に扱うためのデータ管理基盤を構築するものである．特に，LOD の観点からこのようなデータ管理基盤を構築する試みは申請者の知る限り存在しない．データ管理基盤におけるデータ統合の観点において，マルチメディアを含めてオープンデータを統合する試みも本研究の独創性である．さまざまなオープンデータを統合的に管理することで，既存のデータ処理技術の発展をサポートするデータ基盤となる．これらのことから，本研究が目的とする新分野の創造は LOD 研究だけでなくデータを統合的に扱うさまざまな研究や開発，および個々のデータ処理技術の研究開発において重要な分野となる．

【1　研究目的、研究方法など（つづき）】

(3)　本研究で何をどのように、どこまで明らかにしようとするのか

【本申請で取り組む内容】

本研究の目的（図 1）に対し，本申請では基盤技術である (1) データ統合と (2) データ横断検索について研究開発を行う．表 1 にこれらの技術を細分化した要素技術と本申請で扱う内容と年次計画を示す．本申請では，LOD，表形式，画像を対象とし，4 年計画で各要素技術を構築する．

表 1: 本研究の目的達成に必要な要素技術と本申請で取り組む内容および年次計画

技術カテゴリ	要素技術	対象データ対	本申請	計画年度
データ統合	エンティティ同定	LOD ⋮ LOD	✓	1 年目
		LOD ⋮ 表形式	✓	2 年目
		LOD ⋮ 画像	✓	3 年目
		LOD ⋮ 映像		
	データ補強	LOD ⋮ テキスト	✓	1 年目
		画像 ⋮ テキスト	✓	2 年目
		映像 ⋮ テキスト		
データ横断検索	SPARQL 問合せの拡張	LOD ⋮ 構造化テキスト	✓	1 年目
		LOD ⋮ 表形式	✓	2 年目
		LOD ⋮ 画像	✓	3 年目
		LOD ⋮ 映像		
		LOD ⋮ 混合データ	✓	4 年目

《データ統合のためのエンティティ同定技術》 LOD 同士のエンティティ同定 (Entity Alignment; EA) において，既存手法は LOD 間で保有する情報が似ている場合にうまくいくことが示されている．一方で，汎用的な LOD とドメイン特化の LOD のように LOD 間で保有する知識に偏りがある場合に，既存の EA 手法の性能が明らかでない．本研究では，後者の場合における既存手法の性能を明らかにするとともに，知識の偏りに対して頑健な EA 手法を構築する．

テキストに対するエンティティ同定 (Entity Linking; EL) は同じ文脈中に出現するエンティティ間の関係を利用し，LOD 中のエンティティとの同定を行う．表形式データおよび画像において，この「文脈」は自明でないため，何らかの方法で与える必要がある．表形式同士のエンティティ同定 (Entity Resolution; ER) において，エンティティが属性値の組合せで表現されることから，ER ではこの組合せを文脈として扱っている．LOD はグラフ構造を持つため，表形式よりも複雑な構造を有している点で ER よりも複雑な問題である．本研究では，ER 手法を一般化し，グラフ構造を有する LOD と表形式データのエンティティ同定を実現する．一方で，画像に対する物体検出やキャプショニングの技術を活用し，画像中のエンティティに関する記述テキストを導出し，これを文脈としてエンティティ同定を行う．本研究では，この方法に加えて，エンティティ同定のための画像特徴を自動的に学習する End-to-End 方式も開発する．

《データ統合のためのデータ補強技術》 エンティティ同定の際に，各データが保有する情報が異なることが一因となりエンティティ同定を正確に行えないことがある．既存の技術（例えば，[1]）は，あるエンティティについて，一方のデータ中で関連性のあるエンティティは他方のデータ中でも同様の関連性を持つと仮定している．これは，両データが十分な量のデータを持っていることを前提としている．しかし，一般に両データがともに十分な量のデータを持っているとは限らない．この問題に対処するために，エンティティ同定の際に外部知識を取り入れる方法を開発する．各データにおいてエンティティに関連するテキストはエンティティ同定において重要なヒントとなる [2]. しかし，データによってすでに関連付けられているテキストの量や質は異なる．インターネットの普及に伴い多くの情報が Web に発信されるようになり，Web には多様な情報が存在する．その中には，単一のデータには含まれない補足情報も存在しており，上述したエンティティ同定における情報不足を解消することが期待される．本研究では，Web 上の情報を外部知識として活用し，エンティティ同定のための補助データとするためのデータ補強技術を開発する．その際に，Web 上の情報はノイズが多いため，ノイズを考慮した手法の構築が必要となる．また，Web 上のデータにもさまざまなフォーマットのデータが存在しており活用可能性は十分にあるが，本申請ではテキストデータに限定し，他のフォーマットについては後続の研究課題にて取り組む．

【1　研究目的、研究方法など（つづき）】

《横断検索のための SPARQL 問合せ拡張》 オープンデータにはそれぞれ適したフォーマットが存在し，それぞれ適切な処理方法が異なる．そのため，従来ではそれぞれにあった方法でデータにアクセスし，データの処理後に統合していた．このやり方では，データ量が爆発的に増える状況において効率的でない．そこで，検索プロセスを複数のデータ間で共有することで検索を効率化する横断検索技術が必要となる．LOD において，この横断検索は LOD とテキストの間においてさえ，技術が確立していない．そのため，本申請では構造化テキスト，画像の 2種類を対象に LOD との横断検索技術を開発する．具体的には，LOD における標準的な問合せ言語である SPARQL を拡張し，構造化テキストや画像への同時問合せを可能にする．加えて，効率的な検索を実現するために検索結果の共有による問合せ最適化を実現する．

【年次計画】

本研究は，表 1 に示す年次計画で遂行する．以下で，各年度の計画を説明する．

《1 年目》 LOD における知識を補完するために他のドメインの LOD との**エンティティ同定** (Entity Alignment; EA) に取り組む．特に，LOD が持つ知識に偏りがある場合の既存手法の頑健性を明らかにする．そのために，頑健性を評価できるデータセットの構築，既存手法の評価および頑健性を有する EA 手法の構築に取り組む．また，エンティティに対する補助情報をテキストから確保する**データ補強**の方法論を確立し，頑健な EA 手法を構築する．加えて，**横断的な検索**について，LOD 中のエンティティと関連付けられた構造化テキスト (XML, JSON) を対象に SPARQL の拡張およびその処理系を構築する．申請者はこれに関する基礎技術と予備実験を行っており [3]，これを複雑な問合せを扱えるように拡張・効率化し実現する．

《2 年目》 LOD と表形式の**エンティティ同定**において，表形式同士のエンティティ同定 (Entity Resolution; ER) の技術を基本に LOD のグラフ構造を取り入れたエンティティ同定手法を構築する．また，3 年目の LOD と画像の統合に向け，被写体にエンティティを含む画像とそのエンティティの記述を含むテキストを Web から収集し，画像データに対し画像キャプショニングを用いて**データ補強**を行う．加えて，**横断的な検索**について，LOD 中のエンティティと関連付けられた表形式データ (CSV) を対象に SPARQL の拡張およびその処理系を構築する．

《3 年目》 LOD と画像の**エンティティ同定**において，画像キャプショニング技術を活用する手法を構築する．画像キャプショニングは多量の学習データが必要なため，2 年目に収集した画像とテキストを用いて学習データを拡充し，画像と LOD におけるエンティティ同定を実現する．また，画像データを活用した**データ横断検索**を実現するために，画像を入力したエンティティ検索，画像の類似度を条件とするエンティティ検索を SPARQL 問合せの拡張として実現する．

《4 年目》 実応用に向け，統合されたデータおよび横断検索の有用性を向上させる．1 年目および 3 年目に開発した技術をもとに，異種データ複合問合せを実現し，単一の SPARQL 処理系で異種データを検索できるシステムを構築する．加えて，キーワード検索や構造的問合せなどのユーザフレンドリな検索インターフェースを開発し，ユーザビリティを向上させる．

〔参考文献〕

[1] S. Pei et al., “REA: Robust Cross-lingual Entity Alignment Between Knowledge Graphs”, KDD’20, pp.2175-2184, 2020
[2] X. Tang et al., “BERT-INT: A BERT-based Interaction Model For Knowledge Graph Alignment”, IJCAI’20, pp.3174-3180, 2020
[3] T. Komamizu“SPARQL with XQuery-based Filtering”, ISWC’20 Poster, 2020（採択済み）

【研究代表者・分担者の役割】

研究代表者・駒水は，研究全体の統括を行い，各要素技術開発の具体化と遂行の役割を持つ．加えて，データ統合における要素技術に関する研究開発および実装・実験を行う．**研究分担者・波多野**は，構造化テキストに対する検索技術に明るいことから，データ横断検索における要素技術の開発を担当する．**研究分担者・井手**は，画像を中心としたマルチメディアコンテンツ処理に明るいことから，画像を対象とした技術開発を担当する．システム開発や実験については，研究代表者・分担者の研究室の博士後期・前期課程学生を研究補助者として雇用し，適宜分担する．

２　本研究の着想に至った経緯など

本欄には、(1)本研究の着想に至った経緯と準備状況、(2)関連する国内外の研究動向と本研究の位置づけ、について１頁以内で記述すること。

(1)　本研究の着想に至った経緯と準備状況

《本研究の着想に至った経緯》 オープンデータ利活用技術の発展は，データ公開後の再利用性の見通しを立てるために重要である．申請者は，**科学研究費補助金・若手研究 (18K18056)** にて，データ分析の観点からオープンデータ活用に関する研究に取り組んだ．その過程で，テキスト以外の多種多様なオープンデータを活用することで，より高度で多様な分析ができることに気づいた．同時に，異種オープンデータ間の相互利用性の低さにも気づかされた．その要因を調査・検討し，データのフォーマットや内容記述の粒度の違いが相互利用を困難にしているという結論に至った．さらに，オープンデータ活用研究はテキストを基本にしていることが多く，画像や映像などのマルチメディアへの取り組みは限定的である．しかし，マルチメディアは言語化の難しい知識を内包しており，これを取り入れることでオープンデータ活用をさらに発展できると考えた．このような経緯から，申請者は異種オープンデータの統合・活用を行う本研究を着想するに至った．

《準備状況》 本研究における**データ統合**について，初年度に取り組む EA についての 2020 年 9 月時点での研究調査は完了している．それぞれの研究におけるソースコードがまとめられており[2]，すぐに利用できる環境にある．2 年目以降の課題については，初期段階の調査を終えている．研究や開発の速度が早いことから常に調査は必要であるが，本研究遂行するための手法の目星はついている．また，**データ横断検索**については，LOD と構造化テキストを横断的に検索するためのナイーブな手法の構築と予備実験をすでに完了している [1]．3 年目の画像との横断検索については順次調査を進めており，すでに効率的な画像検索手法について目星がついている．以上のように，本研究で取り組む内容に向けた準備は十分にできている．

(2)　関連する国内外の研究動向と本研究の位置づけ

《関連する国内外の研究動向》 本研究に最も関連する国際会議としては，セマンティックウェブに関する会議 ISWC がある．ISWC では，マルチメディアに関する議論はほとんどされず，標準的な LOD における議論が中心となっている．また，LOD とテキストにおける横断的な研究は一時期行われていたものの，特に目立った成果は出ておらず，未だに取り組むべき課題とされている．一方で，データ工学に関する国際会議である SIGMOD, VLDB, ICDE などでは，データ統合に関する議論が行われている．LOD 間のエンティティ同定 (EA) や表形式データ間のエンティティ同定 (ER) についても議論されている [2]．同様の研究は，自然言語処理分野や機械学習分野でも議論されている．これまでの研究で利用されるデータは，DBpedia や YAGO, Wikidata といった多岐にわたる知識を潤沢に蓄えた百科辞典のようなデータが用いられている．これらは知識がお互いに広く十分にあるため，データ統合が比較的容易に行える．そのため，知識の量や幅に違いがあるデータに対して既存手法が十分な性能を発揮できるかは明らかになっていない．さらに，LOD 外部の情報を利用した手法は申請者の知る限り存在しない．

《本研究の位置づけ》 本研究は，エンティティ同定における新たな課題を提起することで，同技術を実用的なレベルに押し上げるための研究である．また，異種データの統合を LOD を中心として行った研究は数少なく，さらに，異種データが統合された LOD の活用に関する研究は他に見られない．これらのことから，見落とされていた重要な研究分野を発掘し，新たな研究領域 Linked Open Multimedia Data Management を開拓する点が本申請の特徴であると言える．

〔参考文献〕

[1] T. Komamizu“SPARQL with XQuery-based Filtering”, ISWC'20 Poster, 2020（採択済み）
[2] Z. Sun et al., “A Benchmarking Study of Embedding-based Entity Alignment for Knowledge Graphs”, Proc. VLDB Endow, Vol.13, Iss.11, pp.2326-2340, 2020

[2] https://github.com/THU-KEG/Entity_Alignment_Papers

３　応募者の研究遂行能力及び研究環境

本欄には応募者（研究代表者、研究分担者）の研究計画の実行可能性を示すため、(1)これまでの研究活動、(2)研究環境（研究遂行に必要な研究施設・設備・研究資料等を含む）について２頁以内で記述すること。
「(1)これまでの研究活動」の記述には、研究活動を中断していた期間がある場合にはその説明などを含めてもよい。

(1)　これまでの研究活動

【研究代表者：駒水 孝裕】

研究代表者は，LOD のデータ管理，構造化データに対する検索，不均衡データ分類に取り組んでおり，本研究遂行のための中心的な知見を有しているため，本研究の代表者として適切である．

- **LOD のデータ管理：**研究代表者は，LOD におけるエンティティの検索 [1, 2, 3, 4, 5] や類似するクラスや述語のクラスタリング [6]，汎用的な分析システム [7, 8] の研究開発を通して，オープンデータを活用する際に技術的な専門知識を必要としないアクセス・分析手法を実現した．[1] では，キーワード検索を用いて LOD 中のエンティティを検索する手法を提案し，ベンチマークにて最良の検索性能を実現した．また，[7, 8] では，LOD 中の数値データを対象に，多次元分析を行うための，データ前処理を効率化した．本申請に関連する研究として，研究代表者が法令 LOD と Wikipedia をもとに作られた LOD である DBpedia に対してエンティティ同定 (EA) を試みた [9]．この際に，本研究の着眼点である，それぞれの情報の偏りによるエンティティ同定の困難さについての着想を得た．
- **構造化データに対する検索：**研究代表者は構造化データである XML データ対象にユーザフレンドリな検索システムを構築した [11, 12]．キーワード検索では XML の持つ構造を捉えきれないため，この研究では構造情報や文書中の単語を用いた対話的な検索システムであるファセット検索を実現した．また，[10] では，構造化データの紐付けられたエンティティに対して，従来では異なる問合せ言語を別々に実行していたものを，単一言語での問合せを可能にした．
- **不均衡データ分類：**研究代表者は機械学習・データマイニングにおける基礎技術の一つであるデータ分類において，クラスラベルごとの事例数に偏りがある不均衡なデータに対するデータ分類に取り組んできた [13, 14, 15]．この不均衡性は現実のあらゆるデータに存在し，分類器が多数派を偏重し分類を誤り易くなるという問題がある．この問題に対し，[14] では不均衡性を有するデータに対して効果的なアンサンブル手法を提案し，[15] でさらなる性能改善を実現した．

〔参考文献〕※下線は研究代表者

[1] T. Komamizu, “Random walk-based entity representation learning and re-ranking for entity search”, Knowl. Inf. Syst., Vol.62, Iss.8, pp.2989-3013, 2020

[2] T. Komamizu, “Graph Analytical Re-ranking for Entity Search”, EYRE@CIKM’18, 2018

[3] T. Komamizu, “Learning Interpretable Entity Representation in Linked Data”, DEXA’18, pp.153-168, 2018

[4] T. Komamizu, T. Amagasa, H. Kitagawa, “CROISSANT: Centralized Relational Interface for Web-scale SPARQL Endpoints”, iiWAS’17, pp.284-288, 2017

[5] T. Komamizu, S. Okumura, T. Amagasa, H. Kitagawa, “FORK: Feedback-aware ObjectRank-based Keyword Search over Linked Data”, AIRS’17, pp.58-70, 2017

[6] T. Komamizu, T. Amagasa, H. Kitagawa, “Interleaving Clustering of Classes and Properties for Disambiguating Linked Data”, ICADL’16, pp.251-256, 2016

[7] T. Komamizu, T. Amagasa, H. Kitagawa, “H-SPOOL: A SPARQL-based ETL Framework for OLAP over Linked Data with Dimension Hierarchy Extraction”, Int. J. Web Inf. Syst., Vol.12, Iss.3, pp.359-378, 2016

[8] T. Komamizu, T. Amagasa, H. Kitagawa, “SPOOL: A SPARQL-based ETL Framework for OLAP over Linked Data”, iiWAS’15, pp.49:1-10, 2015 (**Best Paper Award**)

[9] 駒水 孝裕, 小川 泰弘, 外山 勝彦, “法令沿革 LOD 構築のための DBpedia における法令エンティティの同定”, 第 51 回人工知能学会セマンティックウェブとオントロジー研究会, SIG-SWO-051-06, 2020 年（査読なし）

[10] T. Komamizu“SPARQL with XQuery-based Filtering”, ISWC’20 Poster, 2020（採択済み）

[11] T. Komamizu, T. Amagasa, H. Kitagawa, “Facet-value Extraction Scheme from Textual Contents in XML Data”, Int. J. Web Inf. Syst., Vol.11, Iss.3, pp.270-290, 2015

[12] 駒水 孝裕, 天笠 俊之, 北川 博之, 異種 XML データに対するファセット検索手法の提案, 情報処理学会研究報告 (DD), 2009-DD-073(7), pp. 1-8, 2009 年 9 月（査読なし，**情報処理学会山下記念研究賞**）

[13] T. Komamizu, R. Uehara, Y. Ogawa, K. Toyama “MUEnsemble: Multi-ratio Undersampling-based Ensemble Framework for Imbalanced Data”, DEXA’20, p.213-228, 2020

[14] 植原 リサ, 駒水 孝裕, 小川 泰弘, 外山 勝彦: 弱分類器の調整に基づく不均衡データ向けアンサンブル・フレームワーク, WebDB Forum’19, pp.81-84, 2019（**FUJITSU 賞**, **株式会社 FRONTEO 賞**, **マイクロアド賞**）

[15] 植原 リサ, 駒水 孝裕, 小川 泰弘, 外山 勝彦: 不均衡データ分類フレームワークにおけるサンプリング比率の最適化, DEIM’20, F8-2, 2020（査読なし，**オンラインプレゼンテーション賞**）

【3　応募者の研究遂行能力及び研究環境（つづき）】

【研究分担者：波多野 賢治】

研究分担者・波多野は構造化テキスト検索に関する研究に対する知見を有する [1, 2]．特に，構造化データのひとつである XML データ処理に関する研究をしてきており，本研究における「LOD と構造化テキストの統合と検索」における重要な知見を有している．加えて，テキストからの固有表現抽出と抽出された固有表現からの知識獲得に関する研究も行っている [3]．さらに，グラフデータに対する効率的な検索についても研究しており [4]，データ横断的な検索に必要な知見も有している．これらのことから，同研究分担者は研究代表者への知見の提供や共同研究開発を通して，本研究の目的達成に貢献できるため，本研究の分担者として適切である．

〔参考文献〕※下線は研究分担者

[1] A. Keyaki, J. Miyazaki, K. Hatano, G. Yamamoto, T. Taketomi, H. Kato, “Fast incremental indexing with effective and efficient searching in XML element retrieval”, Int. J. Web Inf. Syst., Vol.9, Iss.2, pp.142-164, 2013

[2] 波多野 賢治, 絹谷 弘子, 吉川 正俊, 植村 俊亮, “XML 文書検索システムにおける文書内容の統計量を利用した検索対象部分文書の決定”, 電子情報通信学会論文誌, Vol.J89-D, No.3, pp422-431, 2006（**平成 18 年度電子情報通信学会論文賞受賞**）

[3] K. Kusu, N. Makino, T. Shioi, K. Hatano, “Calculating Cooking Recipe’s Difficulty based on Cooking Activities”, CEA@IJCAI’17, pp.19-24, 2017

[4] K. Kusu, K. Hatano, “Recurrent Path Index for Efficient Graph Traversal”, BigData’19, pp.6107-6109, 2019

【研究分担者：井手 一郎】

研究分担者・井手は画像を中心とするマルチメディアコンテンツのデータマイニング技術に関する知見を多数有する．本申請に関係する技術としては，画像キャプショニングに関する研究がある [1, 2]．同分担者は，画像キャプショニングに対して，心理学で用いられる画像に対する印象粒度を表す心像性を取り入れ，心像性に合わせたキャプションの生成を実現している [1, 2]．また，本申請で提唱する Linked Multimedia Open Data Management 領域において，今後必要とされる映像に対するデータ工学技術の適用経験を有する [3, 4]．これらのことから，同研究分担者は本申請における画像処理技術に関する研究に対して貢献するとともに，提案領域の発展性についての議論を深めることができることから，本申請における研究分担者として適切である．

〔参考文献〕※下線は研究分担者

[1] K. Umemura, M. A. Kastner, I. Ide, Y. Kawanishi, T. Hirayama, K. Doman, D. Deguchi, H. Murase, “A study on image captioning considering its imageability”, IEICE Tech. Rep. (MVE), Vol.119, Iss.57, pp.165-169, 2020 (**MVE Award**)

[2] Marc A. K., I. Ide, F. Nack, Y. Kawanishi, T. Hirayama, D. Deguchi, H. Murase, “Estimating the imageability of words by mining visual characteristics from crawled image data”, Multim. Tools Appl. Vol.79 Iss.25-26, pp.18167-18199, 2020

[3] F. Nack, I. Ide, “Why did the Prime Minister resign? -Generation of event explanations from large news repositories-”, ACM-MM’11, pp.313-322, 2011

[4] K. Doman, T. Tomita, I. Ide, D. Deguchi, H. Murase, “Event Detection based on Twitter Enthusiasm Degree for Generating a Sports Highlight Video”, ACM Multimedia’14, pp.949-952, 2019

(2)　研究環境

研究代表者・駒水の研究施設には，名古屋大学の居室を利用する．研究設備としてすでに数台の計算機が設置されているが，本研究とは別目的であるため別途購入する．本研究を効率的に遂行するため，高速なディスクアクセスと十分なメモリおよび深層学習のための GPU を搭載した計算機が必要である．具体的には，1TB 以上の SSD と 128GB 以上のメモリ，最新の GPU を搭載したワークステーションを購入する．ソフトウェア開発および実験のためのノート PC やタブレット PC は適宜購入する．**研究分担者・波多野**の研究施設には，同志社大学の波多野研究室の実験室を利用する．研究代表者と同様に，研究遂行のための計算機やノート PC，タブレット PC を購入する．**研究分担者・井手**の研究施設には，名古屋大学の井手研究室の実験室を利用する．研究代表者と同様に，研究遂行のための計算機やノート PC，タブレット PC を購入する．

研究代表者と研究分担者の連携は，オンラインコミュニケーションツールを基本とし，必要に応じて対面での議論を行う．研究成果は研究代表者が集約するため，研究分担者は成果を研究代表者に共有する．連携体制はすでに確立しており，共同で研究を進める準備はできている．

４　人権の保護及び法令等の遵守への対応（公募要領４頁参照）

本欄には、本研究を遂行するに当たって、相手方の同意・協力を必要とする研究、個人情報の取扱いの配慮を必要とする研究、生命倫理・安全対策に対する取組を必要とする研究など指針・法令等（国際共同研究を行う国・地域の指針・法令等を含む）に基づく手続が必要な研究が含まれている場合、講じる対策と措置を、１頁以内で記述すること。
個人情報を伴うアンケート調査・インタビュー調査・行動調査（個人履歴・映像を含む）、提供を受けた試料の使用、ヒト遺伝子解析研究、遺伝子組換え実験、動物実験など、研究機関内外の倫理委員会等における承認手続が必要となる調査・研究・実験などが対象となります。
該当しない場合には、その旨記述すること。

アンケート調査やインタビュー調査は予定されておらず，個人情報を伴うデータを扱わないため，基本的には該当しないが，研究遂行において個人情報や著作権に関わるデータを扱う場合には法令やライセンスに従いデータの利用を適切に行う．

名古屋大学大学院情報学研究科に属する研究代表者・駒水と研究分担者・井手について，被験者を用いた実験が必要となった場合は，名古屋大学大学院情報学研究科に届出を行い，研究科における規定「倫理審査に関係する研究実施計画の取扱いについて」に基づき，審査を受けてからの実施を行うものとする．

研究分担者・波多野は，同志社大学「人を対象とする医学系研究に関する倫理講習」を受講済みである．アンケート調査を含む人を対象とする研究を行う際には，同研究分担者が所属する組織の研究倫理講習を受講した上で，研究倫理委員会に申請書を提出，かつアドバイスを受ける必要があるため，組織的にガイドラインの遵守や必要な承認手続きを遂行できるようになっている．

以上に記す通り，研究代表者および研究分担者は「研究の人権の保護および法令等の遵守」のために必要な知識を備えており，十分に対応できると考える．

５　研究計画最終年度前年度応募を行う場合の記述事項（該当者は必ず記述すること（公募要領２５頁参照））

本欄には、本研究の研究代表者が行っている、令和３(2021)年度が最終年度に当たる継続研究課題の当初研究計画、その研究によって得られた新たな知見等の研究成果を記述するとともに、当該研究の進展を踏まえ、本研究を前年度応募する理由（研究の展開状況、経費の必要性等）を１頁以内で記述すること。
該当しない場合は記述欄を削除することなく、空欄のまま提出すること。

研究経費とその必要性

（金額単位：千円）

年度	設備備品費の明細					消耗品費の明細	
	品名・仕様	設置機関	数量	単価	金額	事項	金額
R3	ワークステーション	名古屋大学	1	2,000	2,000		
R3	データ取得用デスクトップPC	名古屋大学	1	250	250		
R3	検索システム用デスクトップPC	名古屋大学	1	300	300		
R3	ノートPC（研究代表者用）	名古屋大学	1	300	300		
R3	タブレット端末（研究代表者用）	名古屋大学	1	100	100		
R3	開発用デスクトップPC（研究分担者用）	同志社大学	1	500	500		
R3	ノートPC（研究分担者用）	同志社大学	1	300	300		
R3	タブレット端末（研究分担者用）	同志社大学	1	100	100		
R3	開発用デスクトップPC（研究分担者用）	名古屋大学	1	500	500		
R3	ノートPC（研究分担者用）	名古屋大学	1	300	300		
R3	タブレット端末（研究分担者用）	名古屋大学	1	100	100		
R3				計	4,750	計	0
R4	ノートPC（研究補助者用）	名古屋大学	1	300	300	計算機サプライ品	50
R4				計	300	計	50
R5	ワークステーション	名古屋大学	1	2,000	2,000	計算機サプライ品	50
R5	ノートPC（研究補助者用）	名古屋大学	1	300	300		
R5				計	2,300	計	50
R6	ノートPC（研究補助者用）	名古屋大学	1	300	300	計算機サプライ品	50
R6				計	300	計	50

設備備品費、消耗品費の必要性

【設備備品費】本研究は，深層学習を用いた技術開発を行うため，十分な量のメモリとGPUおよびディスクサイズが必要となる．既存手法の実験において，メモリ128GBと（当時）最新のGPUが用いられていたことから，再現実験用に同等のスペックが必要である．この性能を持つワークステーションを1年目と３年目に購入する．これは，3年目から 1年目とは異なる処理（画像処理）を開始し，それらを並列に研究開発するための計算資源が必要なためである．また，オープンデータや画像データの収集用にディスクサイズの大きいデスクトップPC，オープンデータの横断的検索システム構築用にディスクとメモリサイズの大きいデスクトップPCが必要である．これらに加え，研究代表者，分担者，補助者の開発・議論の環境として，ノートPCおよびデスクトップPCが必要となる．

【消耗品】データの受け渡し・保管などに関する消耗品が必要である．

（金額単位：千円）

年度	国内旅費の明細		外国旅費の明細		人件費・謝金の明細		その他の明細	
	事項	金額	事項	金額	事項	金額	事項	金額
R3	研究会参加（東京1日，1人×2回）	100			研究補助者（2人×3ヶ月）	300	論文英語チェック（2件）	200
R3	計	100	計	0	計	300	計	200
R4	研究会参加（東京1日，1人×2回）	100	調査・発表（アジア，1名×2回）	600	研究補助者（2人×5ヶ月）	500	論文英語チェック（2件）	200
R4	研究成果発表（東京1日，1人×2回）	100	調査・発表（北米，1人×2回）	800			国際会議参加費（6回）	600
R4			調査・発表（ヨーロッパ，1人×2回）	800				
R4	計	200	計	2,200	計	500	計	800
R5	研究会参加（東京1日，1人×2回）	100	調査・発表（アジア，1名×2回）	600	研究補助者（2人×6ヶ月）	600	論文英語チェック（3件）	300
R5	研究成果発表（東京1日，1人×2回）	100	調査・発表（北米，1人×2回）	800			国際会議参加費（6回）	600
R5			調査・発表（ヨーロッパ，1人×2回）	800				
R5	計	200	計	2,200	計	600	計	900
R6	研究会参加（東京1日，1人×2回）	100	調査・発表（アジア，1名×2回）	600	研究補助者（2人×6ヶ月）	600	論文英語チェック（4件）	400
R6	研究成果発表（東京1日，1人×2回）	100	調査・発表（北米，1人×2回）	800			国際会議参加費（6回）	600
R6			調査・発表（ヨーロッパ，1人×2回）	800				
R6	計	200	計	2,200	計	600	計	1,000

旅費、人件費・謝金、その他の必要性

【旅費】本研究の提案領域 Linked Open Multimedia Data Management を国際的に広めるために，交際会議での発表が必須である．また，関連する研究の動向調査のために，セマンティックウェブに関する国際会議（主にヨーロッパで開催），データ工学に関する国際会議（主に北米で開催），画像処理に関する国際会議（アジア開催が多い）に参加する．また，国内でも同様の分野の研究者に対しての周知活動が大事であるため，国内の学会にも積極的に参加し，関係研究者との連携を確立する．なお，初年度は新型コロナウイルスの影響により出張が困難であると考えており，落ち着くであろう2021年度末の国内出張のみ申請している．
【人件費】本研究を効率的に遂行するため，研究代表者・分担者が行う研究に対する補助者を雇用する．
【その他】本研究の内容を正確にかつ可読性が高くするために，英語論文のネイティブチェックを行う．また，本研究の国際会議発表のために参加費が必要となる．

研究費の応募・受入等の状況
（１）応募中の研究費

研究者氏名	駒水 孝裕				
資金制度・研究費名（研究期間・配分機関等名）	研究課題名（研究代表者氏名）	役割	令和3年度の研究経費（期間全体の額）	令和3年度エフォート（%）	研究内容の相違点及び他の研究費に加えて本応募研究課題に応募する理由（科研費の研究代表者の場合は、研究期間全体の受入額）
【本応募研究課題】基盤研究(B)（一般） (R3～R6)	異種オープンデータ活用のためのデータ統合・管理基盤の研究開発	代表	3,400 (20,000) (千円)	30	（総額 20,000 千円）
挑戦的研究（萌芽） (R3～R5)	異種食メディアのエンティティ管理に関する研究 (波多野 賢治)	分担	1,000 (5,000) (千円)	5	この研究課題は，食メディアにおける異種データセットの管理をエンティティ同定の観点から実現するものである．エンティティ同定を行う点では本申請と類似するが，本申請では基盤技術開発およびマルチメディアへの発展が主眼であるのに対し，この研究課題は食メディア特有の課題に取り組むものである．本申請とこの研究課題は相乗効果的にそれぞれの成果を向上させる可能性があり，同時に遂行することでそれぞれの研究を昇華させることができる． （総額 - 千円）
			(千円)		
			(千円)		
			(千円)		

（２）受入予定の研究費

資金制度・研究費名（研究期間・配分機関等名）	研究課題名（研究代表者氏名）	役割	令和3年度の研究経費（期間全体の額）	令和3年度エフォート（%）	研究内容の相違点及び他の研究費に加えて本応募研究課題に応募する理由（科研費の研究代表者の場合は、研究期間全体の受入額）
			(千円)		
			(千円)		
			(千円)		
			(千円)		
			(千円)		
（３）その他の活動				65	
合　　計				100 (%)	