

阪口哲男



- 研究室: 研究棟3階、7D312号室(注: 下記参照)
- メール: saka あつと slis.tsukuba.ac.jp
- WWW: <https://www.sakalab.org/>
(テーマ一覧への補足もあるので、必ず見てください)
- 共同研究室: 情報メディアユニオン3階最奥
 - 正式名称「学系共同研究スタジオ3」(私も普段この部屋に)
 - 鈴木伸崇研究室と共同利用
 - 原則: 1人当りPC1台+仮想マシン(サーバ)1台+α
 - 現在、所属メンバーは0名(大学院を含む)。

2019/10/18

主専攻卒研説明会

1

動いてナンボの阪口研

- 情報を扱う上での様々な問題を解決したい
 - 「なんかこんなことでけへんかな?」と考えて、
- 「考えた」手法は本当にいけるのか?
 - 実際に作って、動くかどうかを確認・検証する
 - つまり、「動いてナンボ」
- ただし、「動いてナンボ」は必要条件
 - 単に動かすだけでは十分ではない
 - 動かした上で評価する

2019/10/18

主専攻卒研説明会

2

テーマの方向性(阪口研)

- 情報共有・伝達基盤技術
 - 情報の共有や伝達を快適に行う(大目標)
 - そのために様々なアプリケーションシステムが構築されている
 - 「快適」には「安全」や「安心」も含まれる
- その構築を支える技術の開発が元々の目的
 - 例題としての構築そのものも対象にする
- そのシステムは快適か? 開発上の課題は?
 - そういったことの改善を目指したい

2019/10/18

主専攻卒研説明会

3

テーマの決め方(阪口研)

- 阪口が考えているテーマの一端を選ぶ
 - 工学部的「ノリ」?
 - 世の中における位置づけも意識
 - 学会発表も視野に入れる?
- 持ち込みテーマ
 - 学生さんからの持込
 - ただ、卒研として「もの」になるように議論を重ねるので、最終的には結構変わります
 - 最初に独創に富んだアイデアが欲しい

2019/10/18

主専攻卒研説明会

4

最近の研究トピック

- Crowd4U/FusionCOMPプロジェクトに参画
 - Human-Computation/森嶋教授と共同
 - プラットフォーム/記述言語: CrowdSheet (ICDE2018/demo & CAISE2018)
 - 双葉町プロジェクト (白井教授と共同)
- Linked Open Data/Linked Data関係
 - 目的にあうデータ提供元(リポジトリ)をどう探すか?
 - Web上のデータ提供元自動収集 (情報知識学会2015)
 - データセットのメタデータ作成支援の試み
 - リンクへのCrowdSourcing適用 (情報知識学会2018)

2019/10/18

主専攻卒研説明会

5

直近の修士論文テーマ例

- 2016年度
 - Linked Dataクエリ構築支援のための日本語文に基づくグラフ構造の生成 →途中経過を WebDelorum2016にて公開発表
- 2014年度
 - Web ページとしての類似性を利用したLinked Data リポジトリの自動収集手法 →情報知識学会で発表 (2015年次大会)
- 2012年度
 - Web API ドキュメントからの情報抽出によるプログラムライブラリ作成支援
- 2009年度
 - メール配送系における多様な迷惑メール対策の統合管理手法 →情報処理学会の研究会で発表、学生奨励賞受賞
- 2008年度
 - キリルモンゴル語Webページの縦書きモンゴル語への自動変換システム
- 2005年度
 - 縦書きを含む多言語表示のWebアプリケーション開発用ライブラリ
 - 利用者環境適応型の電子図書館データ形式変換システム
 - 言語に依存しない迷惑メールフィルタの開発
- 2003年度
 - 情報発信組織のためのWeb発信情報アーカイブ構築システム

2019/10/18

主専攻卒研説明会

6

これまでの学生の研究テーマ例

- これまでの主要なもの？(含卒研)
 - Linked Data/LOD関係の各種支援
 - リンクへのCrowdSourcing適用
 - Query構築支援
 - その他
 - HTML5を用いた公開鍵認証
 - Web APIのプログラムライブラリ自動生成
 - RSS記事閲覧での情報推薦
 - 言語非依存型迷惑メール対策(いろいろ)
 - デジタル図書館におけるメタデータ関連(いろいろ)
 - More... (太古にはプログラミング言語開発等も)
- 社会人院生(PhD)/リンクリゾルバのログ分析や複合的な情報資源のLOD化等

2019/10/18

主専攻卒研説明会

7

事例紹介 (スライド抜粋)

- 福島県双葉町震災アーカイブズにおけるクラウドソーシングを活用した国際協調に向けた試み (国際シンポジウム@インドネシア)
- CrowdSheet: An Easy-To-Use One-Stop Tool for Writing and Executing Complex Crowdsourcing (CAISE 2018)
- LODデータセット間のリンクにおけるクラウドソーシング適用の試み (情報知識学会年次大会2018)
- Webページとしての類似性を利用したLined Dataリポジトリの自動収集手法 (情報知識学会年次大会2015)

2019/10/18

主専攻卒研説明会

8

福島県双葉町震災アーカイブズにおけるクラウドソーシングを活用した国際協調に向けた試み

Experiments for International Cooperation Utilizing Crowdsourcing in Earthquake Disaster Archives of Futaba Town at Fukushima Prefecture

阪口哲男 (筑波大学図書館情報メディア系)
SAKAGUCHI Tetsuo (Faculty of Library, Information and Media Studies, University of Tsukuba)
saka@slis.tsukuba.ac.jp

2019/10/18

主専攻卒研説明会

9

福島県双葉町の東日本大震災アーカイブズ

- <http://www.slis.tsukuba.ac.jp/futaba-archives/>
- アーカイブズでは様々な写真を撮影・保存
 - 避難所にあった様々なもの
 - 震災後の双葉町のあちこちの風景



2019/10/18

主専攻卒研説明会

10

写真の公開・利活用のために

- 震災時の経験の記録や被災地の記憶として
- 写っているものはどういうものなのか？
- 風景の中に何が写っているのか？
- それがわかるキーワードや説明が必要
- 国際的に共有→キーワードや説明は日本語や英語だけでなく様々な言語が望ましい

2019/10/18

主専攻卒研説明会

11

現在公開中のタスク

- 写真が1枚表示され、キーワードを入力
 - 既に付与されているキーワードも表示
 - 入力キーワードの言語はプルダウンメニューで選択
 - 日本語、英語、インドネシア語など33言語を指定可能
- 自分の知っている言語でその写真にふさわしいキーワードを入力すればマイクロボランティア1件
- もし、その写真にふさわしいキーワードが思いつかなければ、別の写真に切り替えることもできる
- 画面上のメッセージもボランティアの登録により多言語対応可能 (注: 2019年3月現在一時登録中断中)

2019/10/18

主専攻卒研説明会

12

入力時に指定可能な33言語

コード	Language	言語名	コード	Language	言語名
ar	Arabic	アラビア語	mi	Maori	マオリ語
bg	Bulgarian	ブルガリア語	ms	Malay	マレー語
bn	Bengali	ベンガル語	nl	Netherlandic	オランダ語
de	German	ドイツ語	pl	Polish	ポーランド語
el	Greek	ギリシャ語	pt	Portuguese	ポルトガル語
en	English	英語	ro	Romanian	ルーマニア語
es	Spanish	スペイン語	ru	Russian	ロシア語
fa	Persian	ペルシア語	si	Sinhala	シンハラ語
fi	Finnish	フィンランド語	sv	Swedish	スウェーデン語
fr	French	フランス語	sw	Swahili	スワヒリ語
hi	Hindi	ヒンディー語	ta	Tamil	タミル語
id	Indonesian	インドネシア語	th	Thai	タイ語
it	Italian	イタリア語	tl	Tagalog	タガログ語
ja	Japanese	日本語	tr	Turkish	トルコ語
ko	Korean	朝鮮語	ur	Urdu	ウルドゥー語
lo	Lao	ラーオ語	vi	Vietnamese	ベトナム語
			zh	Chinese	中国語

2019/10/18

主催者研説明会

13

アーカイブズのトップページにリンク



2019/10/18

主催者研説明会

14

写真種類の選択(「もの」か「風景」)



2019/10/18

主催者研説明会

15

ポップアップされる入力画面例



2019/10/18

主催者研説明会

16

CrowdSheet: An Easy-To-Use One-Stop Tool for Writing and Executing Complex Crowdsourcing

Rikuya Suzuki, Tetsuo Sakauchi, Masaki Matsubara
Hiroyuki Kitagawa, Atsuyuki Morishima
University of Tsukuba

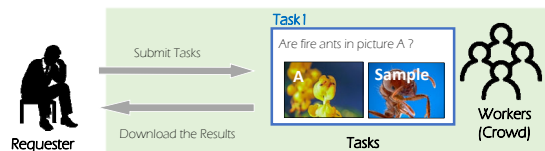
2019/10/18

主催者研説明会

17

Background: Crowdsourcing

Asks people on the Internet to perform tasks



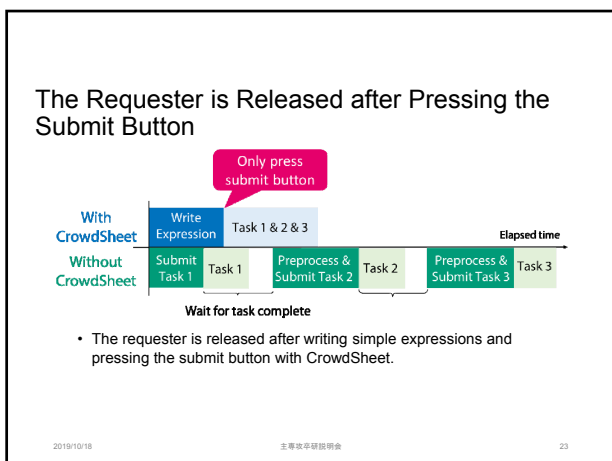
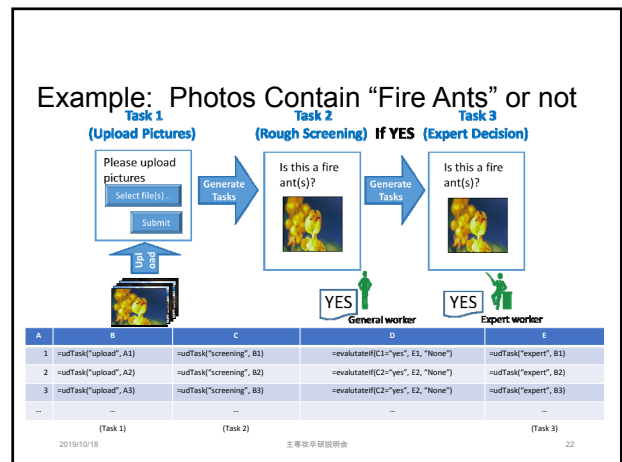
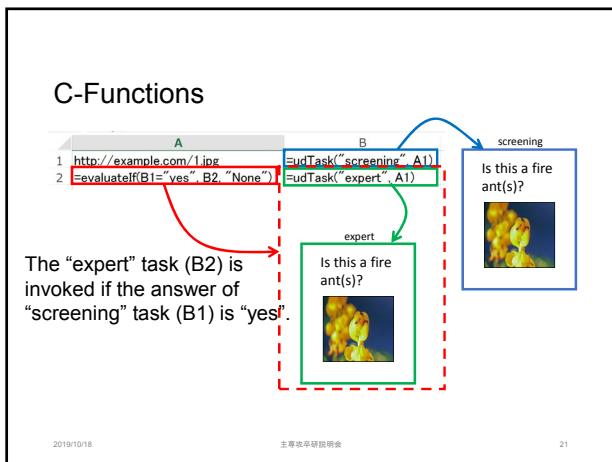
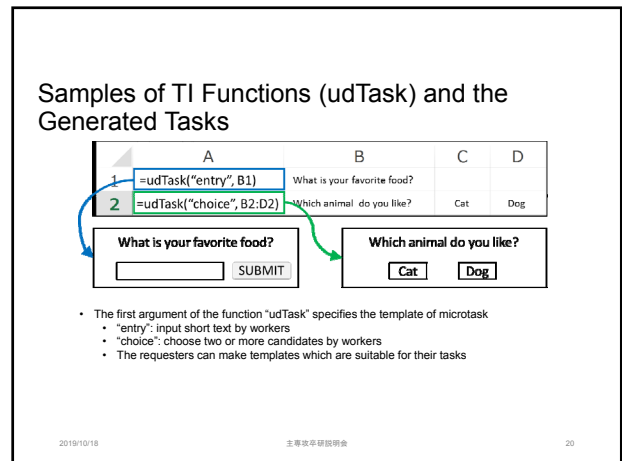
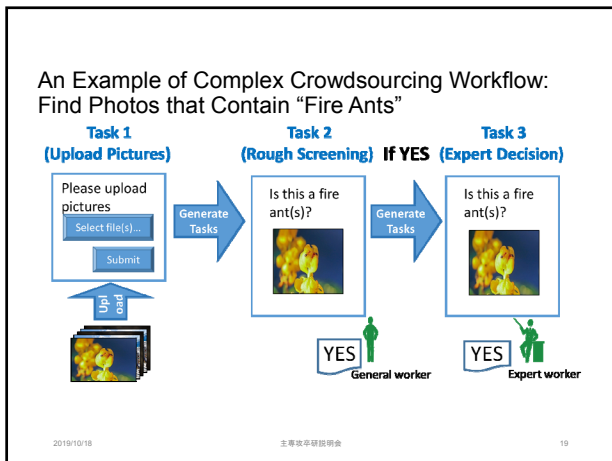
Crowdsourcing Platforms

Help us to register and distribute tasks in an easy way

2019/10/18

主催者研説明会

18



Conclusion

- We proposed CrowdSheet, a spreadsheet for writing and executing complex crowdsourcing applications
- Gave a theoretical result on its expressive power
- Showed that **63% of people** with spreadsheet experiences are able to use CrowdSheet

Future work

- Plan to present the output program in a format closer to that described by humans. This will make it easier to implement additional functions that extend the expressive power of CrowdSheet.

2019/10/18 主催者事務局 24

LODデータセット間のリンク におけるクラウドソーシング 適用の試み

筑波大学
情報学群 知識情報・図書館学類 新井敬樹
図書館情報メディア系 阪口哲男†
†<saka@slis.tsukuba.ac.jp>

2019/10/18 主専攻卒研説明会 25

研究背景 | Linked Open Data(LOD)とは(1/2)

「LODは、様々な情報をコンピュータが処理しやすいデータ形式で公開、データ同士を相互にリンクした、誰でも二次利用できるオープンなデータネットワーク」[1]

LODにおいてリンクは重要

「Linked Data原則」[2]
より多くのものを発見できるように、他のURIへリンクする

「5 star Open Data」[2]
コンテキストを提供するためにデータを他のデータにリンクする

2019/10/18 主専攻卒研説明会 26

研究背景 | リンク作業における問題(2/2)

- ▶リンクを発見・付与してくれるツールがあるが、機械的な処理だけでは、意味を考慮した正確なリンクを付与することは難しい
- ▶そこで人手によるリンク判定をしたいが、それにはデータセット+内の個々のデータを逐一確認する必要がある
- ▶それをデータセットを作成する人だけで網羅性を高められるか？

クラウドソーシングで解決できないか？

†あるサイトで提供されるLODのデータ群をデータセットと呼ぶ

2019/10/18 主専攻卒研説明会 27

研究目的

人手による効率的なリンク判定を実現するために、リンク判定作業にマイクロタスク型クラウドソーシングを適用する手法を開発する。

2019/10/18 主専攻卒研説明会 28

提案手法の概要

リクエストが選択した2つのデータセットからリンク候補を機械的に抽出

↓

リンク候補に関する情報を表形式に変換

↓

表を比較することで、リンク判定をするタスクをクラウドソーシングでワーカに行ってもら

2019/10/18 主専攻卒研説明会 29

評価実験

実験概要

提案手法で作成したタスクが、リンク判定において有効かを検証する。作成したタスク8件を被験者12人行ってもらい、タスクについてのアンケートに答えてもらった。前半4件はURIの短縮形を用いたタスク(TypeA)、後半4件はそのままのURIを用いたタスク(TypeB)である。被験者を3グループに分け、それぞれ別の並び順のタスクを行ってもらった。

タスクに使用したデータセット

- ▶ 鯖江市公共施設(5)
- ▶ 鯖江市避難施設(6)

2019/10/18 主専攻卒研説明会 30

評価実験 | 実際のタスク

表1は「鯖江市の公共施設」、表2は「鯖江市の避難施設」に関するデータの一部分です

TypeA

表1		表2	
属性名	値	属性名	値
rdfs:label	鯖江市立公民館	rdfs:label	避難所
uri:131:name	避難所	uri:131:name	1136.17932
uri:131:lat	0178.2327399	uri:131:lat	1136.17932
uri:131:lon	1136.40027	uri:131:lon	1136.17932
uri:131:address	鯖江市立公民館27番1号	uri:131:address	1136.17932
uri:131:postal	1136.17932	uri:131:postal	1136.17932
uri:131:postal	1136.17932	uri:131:postal	1136.17932
uri:131:postal	1136.17932	uri:131:postal	1136.17932

タスク1-1 「表1が示す場所」と「表2が示す場所」は同じ場所ですか？*

同じ

異なる

わからない

考察

- 行の並べ替えはタスクの行いやすさに影響し、並び順1が最もタスクを行いやすいことがわかった。
- URIを短縮形に変換することで視認性を上げることができることがわかった。
- LOD一般に関する知識とタスクの正答率には相関はほとんどないことがわかった。



提案手法で作成したタスクは、LODデータセットのリンク判定において有効であると考えられる。

課題

- 機械的に、リソースの名前を表すプロパティを抽出すること。
- URIを短縮形に変換する際、そのURIにおいて、より意味のある文字列を抽出すること。
- 表に表示する情報を取捨選択すること。
- 多様なデータセットへの適用性

Webページとしての類似性を利用した Linked Dataリポジトリの自動収集手法

筑波大学
 瀬尾崇一郎 (図書館情報メディア研究科)
 阪口哲男 (図書館情報メディア系/
 知的コミュニティ基盤研究センター)

背景: Linked Open Data

- Webを通じたデータの公開・共有を目指すオープンデータの動きが盛んになってきている
- Linked Dataと呼ばれる方式でオープンデータを公開する Linked Open Data (LOD) が、W3Cによって推奨されている
 - RDFによって構造化し、外部の Linked Data とリンクさせることで、相互連携による利用性を高める
- Linked Dataを提供するWebサイトを、本研究では **Linked Dataリポジトリ(以下、LDリポジトリ)**と呼ぶ

本研究の目的



クローラ型検索エンジンを利用することで、**自動化された機械的な発見手法**による LDリポジトリの収集を行う

データカタログサイト

2019/10/18 主専攻卒研説明会 37

ディレクトリ型による問題点

- データ公開者、または管理者の手によって登録および更新を行う必要がある
 - 登録されないLDリポジトリは発見できない
 - 登録内容の更新がされないとアクセスできなくなる
 - どのデータカタログサイトを使うかわからない
- リポジトリの増加に対応し続けることは難しい

ロボット(クローラ)型の発見方法が必要

2019/10/18 主専攻卒研説明会 38

クローラ型検索エンジンによる自動収集

- クローラ型検索エンジン
 - Webページのリンク関係に沿って収集する検索エンジン
- 一般にWeb上で公開されたLDリポジトリは、ユーザへの告知のために既存のWebページから参照される
 - SPARQL EndpointのWeb UIというWebページに対し、直接リンクを張られることが多い

検索エンジンを利用し、Web UIを取得する

2019/10/18 主専攻卒研説明会 39

自動収集の手がかり: Web UIの類似性

- SPARQL Endpointが持つWeb UIは、構築に利用したツール等の理由から、他のWeb UIと類似することがある

2019/10/18 主専攻卒研説明会 40

提案手法の処理の流れ

2019/10/18 主専攻卒研説明会 41

実験(1/2)

- 提案手法の収集能力はどのぐらいか？
 - 既知のSPARQL Endpointに対する再現率
 - 3種類のWeb UIに対し構築した検索クエリによる検索により、the Datahub中に登録のある該当Web UIを正解セットとした場合の再現率を測定する

	取得により再現できたSE数	the Datahub収録のSE数	再現率
Virtuoso	21	111	0.19
SPARQLer	12	15	0.80
RKB	34	48	0.71

2019/10/18 主専攻卒研説明会 42

実験(2/2)

- 提案手法の収集能力はどのぐらいか？

– 未知のSPARQL Endpointに対する発見能力

- 3種の類似Web UIクラスタそれぞれから抽出したフレーズによって、何件のSPARQL Endpointを取得することができたか
- the Datahubに登録のないSPARQL Endpointが何件あったか

	取得SE数	the Datahub未登録SE数
Virtuoso	59	38
SPARQLer	22	10
RKB	41	7

2019/10/18

主専攻卒研説明会

43

大学院等との関連(阪口研)

- 進学希望者歓迎！(就職希望者もちろん！)
- テーマによっては大学院生にも助言を受けたり、共同テーマになる可能性もあります
 - ただ、今は院生いませんが、、、;-<
- 大学院では他研究室との合同のゼミなども行っています
 - 2019年度: 森嶋研、永森研、三原研

2019/10/18

主専攻卒研説明会

44

まとめ？(阪口研)



- 面談歓迎(メールでの予約推奨)

- 学生いないけど説明会開催！

– 第3回: 10/23(水) 14:00-15:30

– 情報メディアユニオン棟3階

学系共同研究スタジオ3

- 詳細は、

– WWW <https://www.sakalab.org/>

– メール [saka あつと slis.tsukuba.ac.jp](mailto:saka@slis.tsukuba.ac.jp)



2019/10/18

主専攻卒研説明会

45