

NAISTにて取り組みたい研究について

氏名: 新妻 巧朗
試験区分: 情報科学区分
希望研究室: 自然言語処理研究室

1 はじめに

1.1 NAISTで取り組みたいこと

NAISTにて、私が取り組みたい研究テーマは「情報検索システムにおける検索質問に合わせた適切なファセットの推薦、およびファセットの自動生成手法について」である。

このファセットとは、図書館情報学における定義で「あるクラスを2以上の異なる区分特性によって区分したときに得られる下位クラスの総体 [1]」のことを言う。具体的にいうと、検索結果を何らかの区分で絞り込む切り口を提供するものである。

2 研究の概要

2.1 背景・社会的意義

現代社会において情報収集をするためには、検索エンジンを利用することは必要不可欠である。しかし、検索エンジンを適切に活用できず、目的の情報に至れない場面も多い。それは多くの検索エンジンの仕組みが、利用者に対して情報検索能力を要求しているからである。これまでも福島らの研究にて情報検索能力は個人差が大きく、能力差によって情報格差が生じていることが調査されてきた [2]。

こうした課題を解決することで、情報に辿りつけないことで生じる機会損失を減らすことができるのではないかと考えている。

過去に齋藤らによる教育を通して情報検索能力を向上させる研究 [3] も存在しているが、本研究ではシステムの拡張によって解決するアプローチを考えていく。

福島らによって言語能力の高さが情報検索能力の高さに関係しているとわかった [2]。つまり、言語能力の高低が情報検索において、情報格差を生み出していると考えられる。そのため、情報の探索過程で言語能力を要求する場面にて利用者の補助をおこなうシステムを提案したい。

2.2 提案内容

そこで、利用者が入力した検索質問に対して適切なファセットを推薦し、インクリメンタルに検索意図を読み取るシステム (図 1) を提案したい。

情報の探索行動を掘り下げると、検索意図と検索対



図 1: システムのイメージ図

象の意味的な距離を縮めていくプロセスと考えられる。そのため、ファセット検索という手法が役に立つのではないかと考えた。ファセット検索とは、検索システムの利用者に検索対象を何らかの側面で絞り込むファセットを提示し、検索対象を絞り込む検索手法である [4]。これは、システムの利用者が検索意図を言語化する行動をシステムが代行していると言える。そのため、検索エンジンが個人の言語能力に依存している問題にアプローチできると考えている。

3 研究の方法

3.1 従来のファセット検索の課題

ファセット検索の典型的な用例として、Amazon.co.jp [5] の検索結果画面を例にする。ファセット検索は図 2 の画面の左部にあるメニューのようなある分類の検索結果をさらに絞り込む切り口を提供するものである。先に挙げた例では商品データを扱ってい

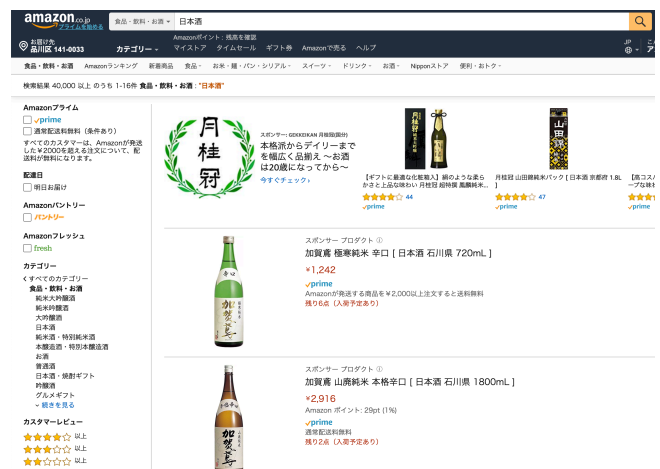


図 2: Amazon の検索結果画面: 日本酒に対するファセット検索

た。このように従来のファセット検索では、索引対象

には既に構造化されたデータを利用することが多く、データを加工することなくファセットを作ることができる。一方で Web のような膨大な数の非構造的な文書を対象とする場合には、二つの課題が生じる。

- 増え続ける非構造的な文書の構造化
- 巨大なファセットの提供方法

3.2 増え続ける非構造的な文書の構造化

特に加工のされていない非構造的な文書群を対象にファセットを作成する場合、文書から切り口となる特徴を取り出して属性データを作成する必要がある。しかし、Web のように文書やその区分特性が変化・増減をし続ける文書であると、人手を利用して属性データを付与するのは現実的でない。そのため、生成可能な属性データを推測してファセットを自動的に生成することで、この課題を解決しようと考えた。

G ファセットは語彙の上位下位のような関係性によって定義される。そのため、実現には索引対象から語彙間の関係性を抽出することが必要である。

3.3 巨大なファセットの提供方法

膨大な量のファセットをそのまま提示したのでは、かえって利用者の負担が増えてしまうためである。Web のような膨大な数の文書が存在する空間を対象とした場合、文書の種類も無数に存在する。その結果、文書を分類することで作成されるファセットの数も同様に膨大な数になり、利用者が絞り込める数ではなくなってしまう。そのため、検索質問に合わせて関連したファセットを推薦することで解決をしたいと考えた。つまり、検索質問から関係するファセットを得る必要がある。検索質問は文書内のキーワードと一致することを念頭に入力されたものであると考えられるため、検索質問は文書内の語彙と同一視することができる。そこから、検索質問とファセットの関係性は、索引対象の語彙間の関係性を抽出することで擬似的に取得できると考えられる。

上記の二つの課題の解決手段として、ともに語彙間の関係性を抽出することの必要性を述べた。これまで語彙同士の関係性を抽出する研究は、情報抽出の領域における banko らによる研究 [6] を皮切りに、OpenIE(Open Information Extraction) という分野にて研究されてきた歴史がある [7]。これらの研究成果を活用して、先に挙げた課題に対応する手段を検討していきたい。

4 これまでの修学経験等

学生の頃は地方の産業構造に関する実証分析について研究してきた。特に卒業研究では総生産と地域を構成する産業に着眼し、経済格差が生じる要因について

分析をしてきた。その後、ソフトウェアエンジニアとして Web サービスに携わる中で、検索システムの利用者が得たい情報をどのように探索しているのか、どうあると嬉しいのかについて考えてきた。特に現在携わっているアルバイト求人のデータベースメディアでは、サイトの利便性の向上のために、どのようにファセットナビゲーションを実現するとよいか、求人検索機能のファセットナビゲーションをどのように実装すべきかなどを試行錯誤する機会に恵まれた。こうした経験が本研究には役立つのではないかと考えている。

5 まとめ

ここまで私が NAIST にて取り組みたい研究テーマについて述べた。

参考文献

- [1] 日本図書館情報学会用語辞典編集委員会編 (2013), 図書館情報学用語辞典 第 4 版
- [2] 福島健介・小原 格・須原慎太郎・生田 茂 (2005), インターネット検索能力の差異に及ぼす 要因の検討 その 1, コンピュータ&エデュケーション VOL.18 2005
- [3] 齋藤ひとみ・三輪和久 (2004), Web 情報検索におけるリフレクションの支援, 人工知能学会論文誌 19 巻 4 号 C (2004 年)
- [4] Daniel Tunkelang (2009), Faceted Search (Synthesis Lectures on Information Concepts, Retrieval, and Services), pp. 21–26
- [5] Amazon.co.jp (最終閲覧日: 2019 年 5 月 23 日), <https://www.amazon.co.jp/>
- [6] Michele Banko, Michael J Cafarella, Stephen Soderland, Matt Broadhead and Oren Etzioni (2007), Open Information Extraction from the Web, IJCAI'07 Proceedings of the 20th international joint conference on Artificial intelligence, pp. 2670–2676
- [7] Christina Niklaus, Matthias Cetto, Andre Freitas, Siegfried Handschu (2018), A Survey on Open Information Extraction, Proceedings of the 27th International Conference on Computational Linguistics