

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て検索質問に合わせた適切なファセットの生成・推薦手法について」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、検索意図と検索対象の意味的な距離を縮めていくプロセスと考えられる。



図 1: システムのイメージ図

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、検索対象を絞り込んでいく検索手法である [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 の画面中左部のメニューのように、ある分類に関する検索結果をさらに絞り込む切り口を提供する。先に挙げた例では商品データを扱っていた。こ

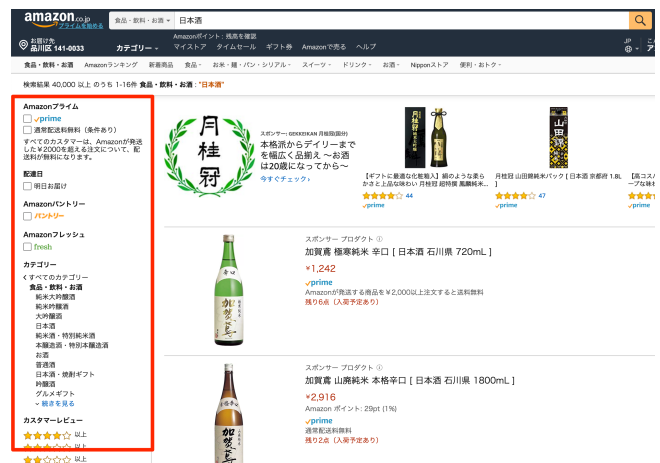


図 2: Amazon の検索結果画面: 日本酒に対するファセット検索

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ているデータを利用することが多く、属

性データを抽出することなくファセットを作成できる。一方で、Web のような膨大な数の文書と無数の区分特性が存在しうる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
- 検索質問に合わせたファセットの提供方法

3.1.1 増え続ける非構造的な文書の構造化

特に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区分を抽出して属性データを作成する必要がある。しかし、Web のように索引される文書が頻繁に増減し、その文書の分類がもろく変化しやすい領域では、人手で文書に属性データを作成していくのは現実的でない。そのため、生成可能な属性データを推測して自動的に生成することでファセットを作成することで解決したい。このとき、ある単語を起点にしたとき、その単語と類似の関係にある他の単語群を束ねて、属性データにすることでファセットを生成できると考えている。例えば、「日本酒」という単語を起点にした場合、「一ノ蔵」「獺祭」は下位語の関係性にあるが、これを束ねることで「銘柄」という概念になる。また、「甘い」「辛い」は修飾の関係にある語だが、これは束ねることで「味」という概念になる。この概念がファセットの区分となり、束ねられた語彙が属性データの値にできると思われる。

3.1.2 検索質問に合わせたファセットの提供方法

従来のファセット検索では、ファセットを提供できるカテゴリが有限で、かつそのファセットも既定のものとして提供される。それは開発段階で提供するファセットを決める必要があるためである。これは、従来では特定の領域に絞った情報検索システムで利用されてきたため、大きな問題となることはなかった。しかし、上記の「非構造的な文書の構造化」の課題が解決され、Web にファセット検索を利用できるようになると課題が生じる。それは、有限でないカテゴリに対してファセットを提供する必要があるからだ。なぜならば、Web を対象とした検索ではファセット検索のカテゴリにあたるものは検索質問になるためだ。そこで、検索質問に最適なファセットを絞り込んで推薦する方法を提案したいと考えた。ファセットは上位クラスと下位クラスの語彙の間に、一对多の意味上の関係を作るものだともみなせる。そのため、検索質問の語彙と上位クラスの語彙の意味が近いファセットを選ぶことで最適なファセットを得られると考えられる。

3.2 課題解決の方向性

二つの課題へのアプローチ方法には、ともに検索対象の語彙間の関係性を抽出することが重要であること

を述べた。これまで語彙間の関係性を抽出する研究は、RE (Relation Extraction) と呼ばれる分野で研究されている。RRE は文中にある二つの Entity を繋ぐ関係を抽出する技術の領域である。これらの研究成果をベースに上記で上げた課題を解決する手法を検討していきたい。

4 これまでの修学経験等

学部で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た。また、社会人ではソフトウェアエンジニアとして Web サービスに携わり、検索システムの利用者が得たい情報をどう探索している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、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検索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では役立つのではないかと考えている。

5 最後に

ここまで NAIST にて取り組みたい研究テーマや自身の経験について述べてきた。私が NAIST を志望するのは、異なるバックグラウンドを持った人間を受け入れるサポート体制が整っており、かつ優れた研究成果を出している大学院であるからだ。こうした NAIST の整った教育・研究環境を活かして、自然言語処理や情報検索の分野に貢献していきたいと考えている。

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/>