

Amazon WEB サービスを用いた効率的な書籍情報獲得システムの開発

工藤 草平[†] 木村 昌臣[†]

芝浦工業大学工学部情報工学科[†]

1. はじめに

近年、インターネット上で本を探し、購入することができるようになってきている。キーワードで検索することにより該当する本の一覧を見ることができる。しかし、「オブジェクト指向でなぜつくるのか」という本は、Java を勉強する上で必要となるかもしれないが、検索キーワード「Java」では検索結果として表示されない。さらに、「Java」で検索すると「やさしい Java」というように探したい内容の本が出てくるだけでなく、「JavaScript」など「Java」という文字を含むが違う内容の本も検索結果として表示されてしまう。また、Java を勉強したいユーザーでも、初心者と上級者では求める本は違うが、書評などの本の詳細情報を見るまでは本の詳しい内容は分からず、本の選択に時間がかかる。このように、インターネット上で欲しい本を探すには、まだまだ不便な点が多い。

そこで、本研究では以上の問題点を解決し、より簡単に求めている本を探すことができるシステムの開発を目的とする。

2. システムの提案

1. で挙げた問題点を解決するためには、システムに以下 2 つの機能が必要である。

1 つ目は、関連のある本が出てこず、違う内容の本が出てきてしまうという問題を解決するために、調べたい内容に関係のある本を効率的に発見できる機能を実現する。そのために本研究では、Amazon[1]の「この商品を買った人はこんな商品も買っています」という Amazon が協調フィルタリングによって提示している書籍情報に着目した。この項目を見れば、キーワード検索では出てこないが、関連性のある本も取り出すことができると考えられる。そこで、今回開発するシステムでは、この本同士の関連性をグラフを用いて可視化する。可視化することで、関連のある本同士がリンクで繋がれ、クラスタ化し、そのクラスタ内の本を見ることで探している内容の本を効率よく見ていくことができる。また、多くのノードからリンクされているノードが人気のある本だと一目で分かるようになる。

本の名称、著者、他の本との関連性等についての情報は「Amazon WEB サービス」を用いて取得することにした。これは Amazon 側で用意しているパラメータを組み合わせた URL を用いて、本の情報だけでなく、カスタマーレビューなども Amazon から受け取れるというものである。

また、2 つ目の機能として、本の詳細を見るまで内容が分からないという問題を解決するために、どんな人向きなのか等、本を買う上で参考となる情報が一目で分かるように表示させる機能を実現する。これは、本のカスタマーレビューやタイトル内に、「やさしい」や「Java」など、その本の難易度を表す単語や内容を表す特徴語の有無をチェックする。そして、特徴語が存在する本には、「初心者向け」や「Java」といった情報を付加する。この情報をグラフ内の本のノードの色や形を変えることでユーザーに提示し、一目で本の内容を特徴付ける情報が分かるようにする[2]。

以下に、本システムの処理の流れを大まかに説明する。

1. 探したい本のキーワードを入力してもらう。
2. 入力されたキーワードでの検索結果を Amazon から取得し、ユーザーに提示する。
3. その中からユーザーに自分が探している本にもっとも近い本を選んでもらう。
4. 選ばれた本と協調フィルタリングによってリンクされている本の情報を取り出す。この本同士のリンクを一定回数辿り、ユーザーにリンク構造を示す(図 1)。

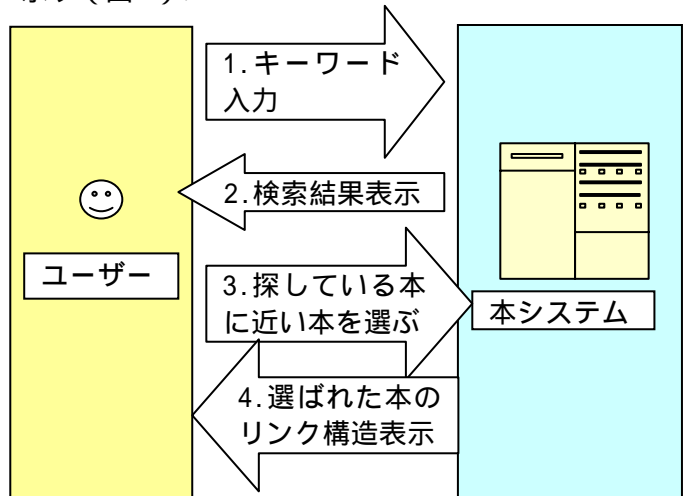


図 1 システム使用時の流れ

Development of efficient book information acquisition system using Amazon WEB service
 Souhei Kudou[†], Masaomi Kimura[†]
 Shibaura Institute of Technology[†]

3. リンク構造の可視化手法

本研究では、本同士のリンク構造を可視化するために、Kamada-Kawai 法を用いる。Kamada-Kawai 法とは、グラフのノード同士を繋いでいるエッジをバネと仮定して、グラフ全体のバネエネルギー（E）が最小となるようにノードの配置を決める手法である。Eは、式（1）の通りである。

$$E = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{1}{2} k_{v_i v_j} (d_{v_i v_j} - l_{v_i v_j})^2 \quad (1)$$

kはバネ定数、lはバネの自然長、dはノード V_i, V_j 間の距離である。

Eが最小となるノードの座標を求めることによって、リンクで繋がっているノード同士が近くに配置され、本の特性毎にクラスタができ、ノードの重ならない見やすい可視化が可能となる。

Eが最小となるノードの座標を求める際に、最急降下法を用いた。しかし、最急降下法では、解が局所解におちいるという問題がある。局所解におちいるとノードが密集した場所に配置され、見やすい可視化ができない。この解決策として、ノードの座標計算処理の途中に、ノードの位置を大きく移動させる処理を加えた。移動後のEが大きき時には移動を取り消し、Eが小さくなれば移動後の座標で計算を続けることで、局所解に嵌ったままノードの座標を計算していくことがなくなった。また、エッジで繋がっていないノード同士には斥力がかかるようにし、葉ノードと繋がっているエッジの長さを短くするなどして、ノード同士が重ならない、より見やすい可視化となるようにした。

4. システムの実装と検証

図2は、本システムがユーザーに提示する本のリンク構造出力画面である。ノードをクリックすると、画面右側にその本の詳細情報が出力される。

図2のの部分には、各本の持つ特徴語が表示される。特徴語は、各本のタイトルとカスタマーレビュー内の単語の出現頻度を用いて抽出する。この中の語が選ばれると、その語を特徴語として持つ本のノードの色が変わる。これにより、本の詳細な情報を見ずに自分の探している内容の本を見つけることができる。

の部分には、「初心者向け」など本を分類するボタンを設けた。本システムでは、本のタイトルやカスタマーレビュー内に、あらかじめ設定しておいた単語（「初心者向け」の項目なら「基本」や「入門」など）が存在した場合、その本に「初心者向け」といったように情報を付加する。

内のボタンを押すと、そのボタンの情報を持った本のノードの形が変わる。これによって、素早

く初心者向け候補の本だけを見ることができる。

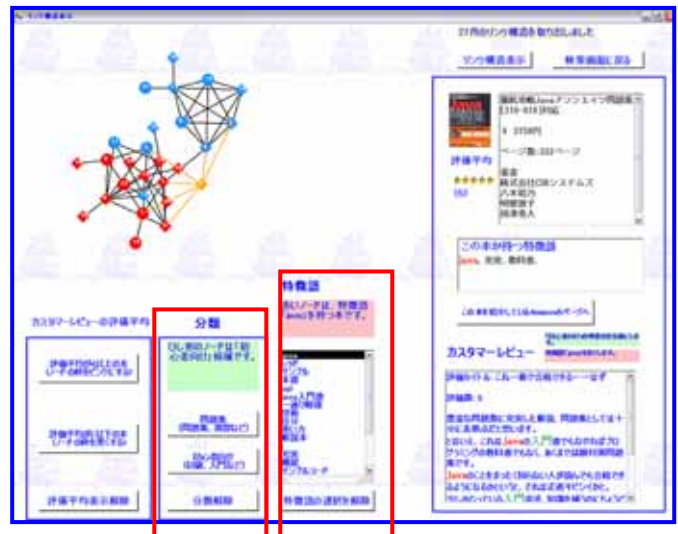


図2 本のリンク構造出力画面

システムを実行した結果、検索ワード「Java」から、オブジェクト指向関係の、キーワード検索では出てこないが、関連性のある本も取り出せた。また、特徴語としても、「オブジェクト指向」という単語が取り出すことができ、関連のある本の探索に役立てることができた。また、同じ特徴語を持ったノード同士は、クラスタ化する傾向も見ることができた。

で実際に「初心者向け」と分類された本20冊を見た結果、5冊が初心者向けの本とは思われないものだった。これは、カスタマーレビュー内に「初心者には難しい」という文章がある本も初心者向けとシステムが判断してしまうため分類対象ではない本も取り出されてしまった。

5. おわりに

本研究では、本同士のリンク構造を可視化、本の特徴語の抽出、本の分類分けをし、本探索を支援するシステムの構築を行った。これにより、本のリンク構造やノードの色や形を見るだけで自分の探している内容の本を効率良く探索できるようになった。今後の課題として、本の分類の精度の向上などが挙げられる。

<参考文献>

- [1] Amazon <http://www.amazon.co.jp/>
- [2] 石川, 木村 : BLOG のトラックバック構造における評価の可視化, W12, 2006-86, pp. 103-104 (2006)
- [3] Amazon Web サービスについて ~ 調査メモ http://www.dap.ndl.go.jp/docs/dap/amazon/amazon_research.html
- [4] 西尾 元宏 : 板ばねモデルを用いたインタラクティブな曲線グラフ描画手法とその応用に関する研究; 東京大学大学院工学系研究科 平成 15 年修士論文