

各言語版 Wikipedia の比較による大学情報整備実態の分析

都市計画専攻 20111246 後閑 晃司

指導教員：渡辺 俊 教授

1. 研究の背景と目的

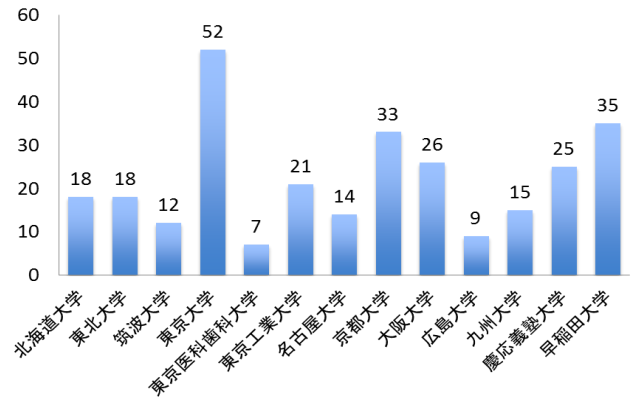
インターネットを通じた情報収集手段の中でも代表的なものに、フリーオンライン百科事典「Wikipedia」が挙げられる。Wikipedia は、ウィキメディア財団(米, サンフランシスコ)により運営され、原則としてすべての人が閲覧・編集を行うことができる。2014年11月16日現在、288の言語版が公開されており[1]、世界中の人々が情報収集を行う上で極めて影響力が大きい。一方で、高等教育における国際競争に関して、「成長戦略第2弾スピーチ(平成25年5月17日)」の中で、「ビッグ・データの活用」「日本の大学を『世界の大学』へ」といった方針が国策として打ち出されている[2]など、ビッグデータをどのように活用するかに注目が集まっているのが現状である。こうした中、ビッグデータのような集合知に映される潜在的な大学のイメージを把握することは、筑波大学の情報発信の指針を検討するにあたり大きな意義があると言える。国内の研究で、Wikipedia の記事編集をコミュニティによる活動として分析したものは比較的少なく、ビッグデータ源としての Wikipedia から大学のイメージ把握を試みているものは見当たらない。当研究では、Wikipedia のビッグデータから、筑波大学の今後の情報発信戦略に関する知見を得ることを目的とする。

2. 方法

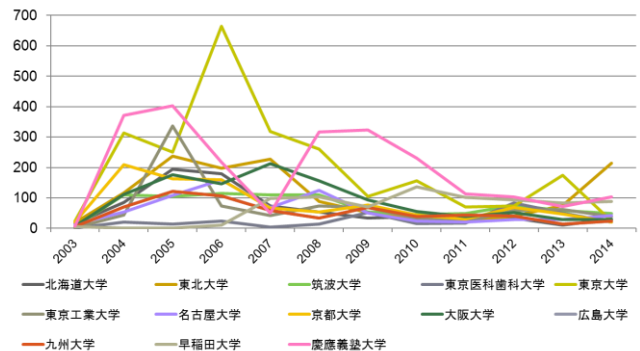
当研究で用いた方法は以下の2つである。(1) Wikipedia の大学に関する記事を対象とした編集履歴の解析。記事の情報が整備される過程で「版」として蓄積されたデータを解析することで、大学イメージの把握を試みる。(2) 学生を対象としたアンケート調査。筑波大学の学生を対象とし、学生による Wikipedia 利用の実態を調べた。以上(1)(2)の解析・調査を通し、Wikipedia の記事に掲載されている情報がいかに整備されているか、また、そうした情報が学生によりいかに利用されているかを明らかにし、筑波大学における今後の情報戦略について考察を加えた。

3. 大学に関する Wikipedia 記事の解析

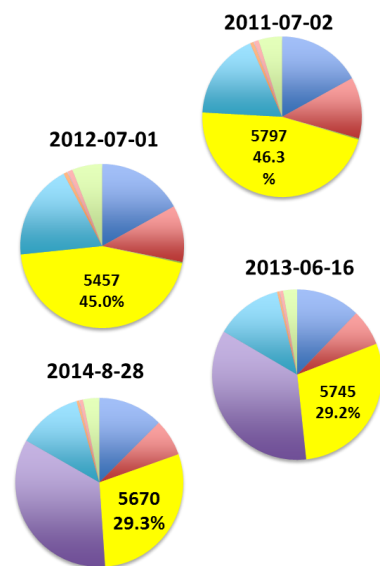
大学に関する Wikipedia 記事の解析では、文部科学省による「スーパーグローバル大学創成支援」制度における【タイプA: トップ型】13校の Wikipedia 記事を対象とした。13校の名前、および Wikipedia のいくつかの言語版で記事が作成されているかを(図1)に、蓄積された版の数(編集の回数と見なせる)の年次的な推移を(図2)に、筑波大学に関する英語版記事の章立ての変化を(図3)に示す。



(図1) 大学別言語版記事作成数



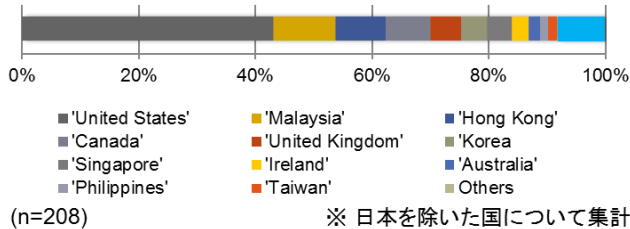
(図2) 日本語版各記事版数年次推移グラフ(2003~2014)



(図3) 「University of Tsukuba」(英) 章立ての年次的変化

(図3)において、学術的なランキングが占める部分には文字数とパーセンテージを付した。学術的なランキングに関する記述は年々増えていることが読み取れる。

続いて、編集履歴に記録された匿名利用者の IP アドレスを基に、編集者の位置情報を取得した。一例として、早稲田大学に関する英語版記事の結果を(図4)に示す。

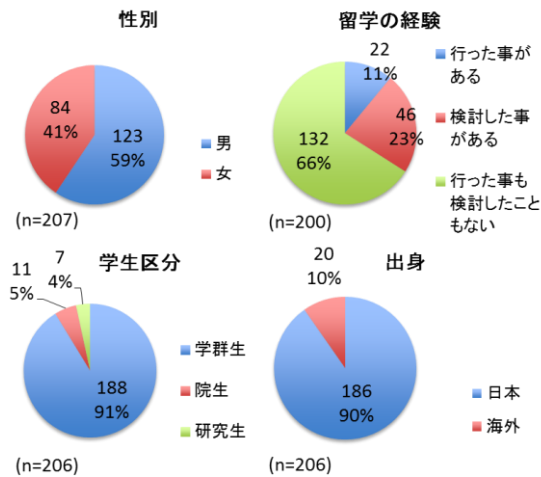


(図4) 「Waseda University」(英) 匿名編集者の国情報

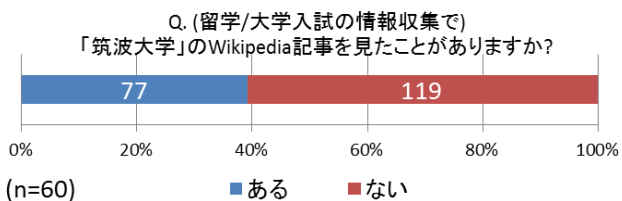
英語版の記事では学術ランキングに関する記述サイズの割合が大きい傾向にあり、英語版の編集者がこれらの学術ランキングを重視していることを示唆する。また、「Waseda University」(英)の記事では、他大学の記事と比較して匿名利用者が世界中に広く分布する傾向にある。匿名編集者の分布は、その地域に大学への興味関心をもつ者がいる、ということを示すという見方もできる。

4. 筑波大学の学生を対象としたアンケート調査

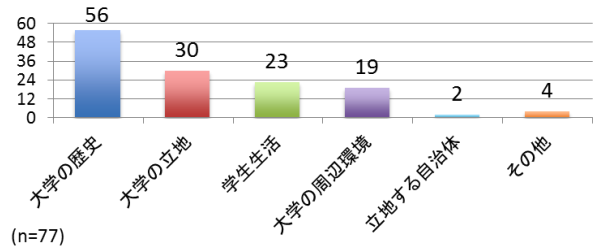
学生による Wikipedia 利用の実態を知るため、学生を対象としたアンケート調査を実施した。回答者属性を(図5)に、筑波大学に関する回答の結果を(図6)(図7)(図8)に示す。



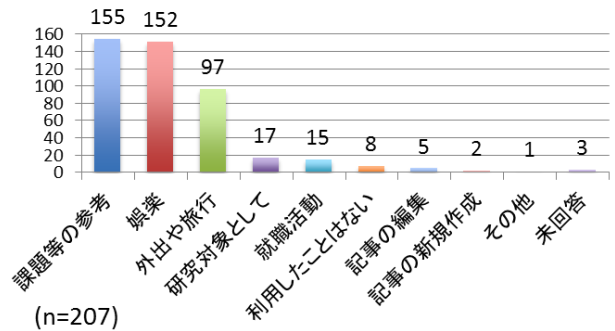
(図5) アンケート回答者属性



(図6) 「筑波大学」に関する Wikipedia 記事の閲覧経験



(図7) 筑波大学に関する Wikipedia 記事から得た情報



(図8) その他の情報収集での Wikipedia の利用目的

5. まとめ

当研究の成果は3点に要約できる。①【タイプA:トップ型】の13校に関する記事の比較を通して、英語版では日本語版と異なる情報整備の基準が採用されていることを明らかにした。② 英語版の大学に関する記事について、編集者の分布に偏りが見られることを明らかにした。③ アンケート調査から、学生は Wikipedia を日常的に利用していることや、その際に得る情報の傾向を明らかにした。以上から、Wikipedia の記事における大学情報整備の実態の把握、および学生にとっての情報収集手段としての Wikipedia の利用状況を明らかにし、大学の国際的な情報戦略に対する知見を得ることができた。

6. 今後の課題

当研究では、記事の本文中の単語については考慮していない。今後は位置情報と見出し内のテキストデータを紐付け、ある地点にいる編集者が記事中にどのような情報(単語)を追加しているかといった解析を行いたい。

7. 参考文献

Web ページは何れも 2015. 2. 3 pm17:00 最終閲覧

[1] Wikipedia: ウィキペディアについて - ja.wikipedia.org/wiki/Wikipedia: ウィキペディアについて

[2] 成長戦略第2弾スピーチ - www.kantei.go.jp/jp/96_abe/statement/2013/0517speech.html

[3] 新井嘉章, 福原知宏, 増田英孝, 中川裕志. Wikipedia を用いた多言語情報アクセスに関する研究: 言語間リンクの分析と応用 (2008, 人工知能学会研究会資料)