

# サイバースペースにおける国内大学のプレゼンスの定量化 Wikipedia 記事の「重要度」測定を通して

○後閑 晃司\*1 渡辺 俊\*2

キーワード：Wikipedia ビッグデータ 大学 中心性指標

## 1. 研究の背景

### 1-1. サイバースペースにおける政策

昨今では、インターネットを通じて、物理的拘束に捉われないこと様々な情報を取得することができる。日本の国策の中に着目すると、グローバル化に特化する大学を採択し、支援を行う「スーパーグローバル大学創成支援制度」<sup>1)</sup>が実施されているほか、「ビッグ・データの活用」「日本の大学を『世界の大学』へ」といった方針が強調されており<sup>2)</sup>、大学の存在感を国際的にアピールすることや、実空間上のみならず、サイバースペース上での政策を検討することが重要となってきたのが現状だと言える。

### 1-2. Wikipedia とそのデータの特徴

サイバースペース上での主要な情報収集手段の一つに、フリーオンライン百科事典「Wikipedia」がある。Wikipedia は、内容の充実した web 上の百科事典を目指すプロジェクトで、様々な事柄に関する情報がボランティアの手により編集されており、2015 年 10 月 1 日現在、**291** の言語で閲覧が可能である<sup>3)</sup>。Wikipedia 内の全てのページは「記事」として扱われ、閲覧や編集は、原則として誰でも自由に行うことができる。記事における編集は「版」という単位で逐一記録されており、ブラウザ上での確認や、XML 形式でのダウンロードも可能である。



図1 Wikipedia 記事の「履歴表示」ページ

## 2. Wikipedia 利用に関する解析と調査

当研究では、先述の制度<sup>1)</sup>における【タイプA：トップ型】13校を対象とし、記事編集についての解析として13校に関する Wikipedia 記事の編集履歴データの解析と、記事の閲覧についての調査として Wikipedia に関する学生アンケートを行ってきた。

### 2-1. Wikipedia 編集の概況

記事編集に関する解析では、前述の編集履歴データを対象としており、独自で作成した Python のプログラムを用いている。【タイプA：トップ型】13校それぞれの大学に関する Wikipedia 記事がいくつかの言語で作成されているかを図2に、日本語版の各校に関する記事の編集回数がどのように推移してきたかを図3にそれぞれ示す。大学ごとに言語版作成数や編集回数には差があることが読み取れる。このことから、Wikipedia の記事を編集する者の興味や態度は、その記事がどの大学を取り扱うかによって大きく異なっていると見える。

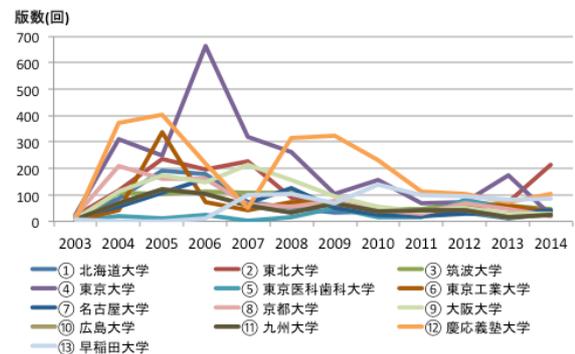


図2 日本語版各記事・版数年次推移グラフ(2003-2014)

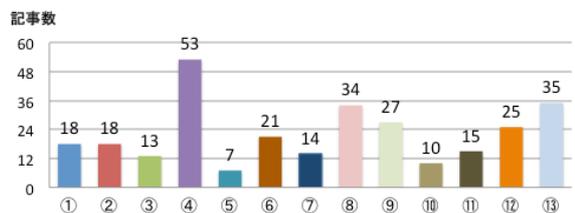


図3 大学に関する各言語版記事作成数(2015. 2. 26 時点)

### 2-2. アンケート調査

Wikipedia がどのように閲覧されているかを把握するため、学生を対象としたアンケートを行った。対象としたのは筑波大学の学生で、合計 207 部を回収している。Wikipedia の日常的な利用や、大学入試、留学といった場面での情報収集について問う内容とした。アンケートの概要を表1に、回答者属性を図4にそれぞれ示す。

Wikipedia の日常的な利用について問う設問では、207 人すべての回答者がなんらかの形で Wikipedia を閲覧したことがあると回答している。また、留学の際の情報収集に関する回答結果をそれぞれ図 5、図 6 に示す。Wikipedia を日常的に閲覧する学生は多く、また大学入試や留学といった場面でも情報収集のために閲覧する学生は一定数存在しており、サイバースペース上で Wikipedia が与える影響力は大きいと言える。

表 1 アンケート実施概要

期間	2014 年 11 月 ~ 2015 年 1 月
対象	筑波大学の学生(学類生, 大学院生, 研究生)
回収数	207 部
実施状況	講義の休み時間、各研究室ゼミ終了後の時間

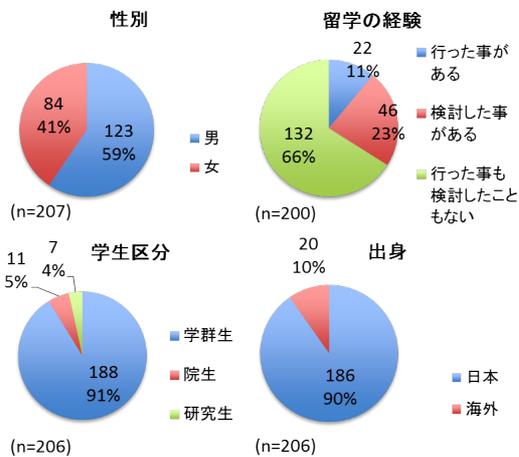


図 4 アンケート回答者属性

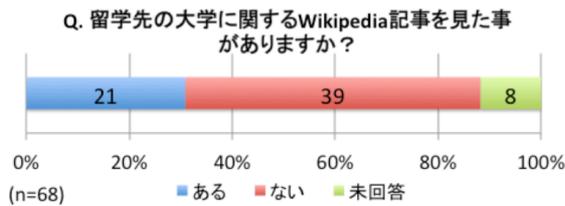


図 5 留学先の大学に関する Wikipedia 閲覧経験

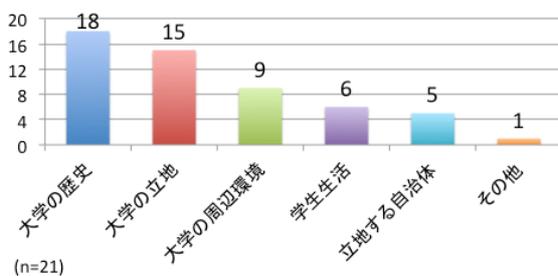


図 6 留学先の大学に関する Wikipedia から得た情報

### 3. Wikipedia に関する中心性指標の測定

Nakayama ら(2007)<sup>4)</sup>は、Wikipedia のデータを用いたシソーラスの新しい構築手法を提案しており、Wikipedia のコーパス的な特徴として URL の一位性、豊富なリンク関係の二つを挙げている。このリンク構造を利用した研究は多く、近年のものとしてはミラノ大学の Laboratory for Web Algorithmics による The Open Wikipedia Ranking(2015. 以下、Wikipedia Ranking)<sup>5)</sup>がある。これは、Wikipedia のリンク構造やアクセス履歴に基づき、ある分野に関するキーワードの”重要度”の測定を試みるプロジェクトである。なお、対応しているのは英語版 Wikipedia のみである。重要度は、Harmonic Centrality, Indegree, Page views, PageRank という 4 つの指標で測定され、ブラウザ上で手軽に閲覧できる。

#### 3-1. Harmonic Centrality (HC)

Yannick(2009)<sup>6)</sup>による、グラフ理論に基づく中心性指標の一つで、ある中心ノードとの距離が n ホップ以内であるノードについて、それぞれの距離(ホップ数)の逆数を重み付けし、それらの総和を中心ノードのスコアとする手法である。Indegree は HC 値を算出するプロセスで、中心ノードから 1 ホップの距離にあるノードのみを考慮した場合のスコアである。なお、Yannick(2009)の定義ではエッジの方向は考慮していないが、Wikipedia Ranking においては他の記事から中心ノードへ向かうアンカーテキストのみをエッジと見なしている。Wikipedia Ranking における Harmonic Centrality と Indegree の重み付けのイメージを図 7 に示す。

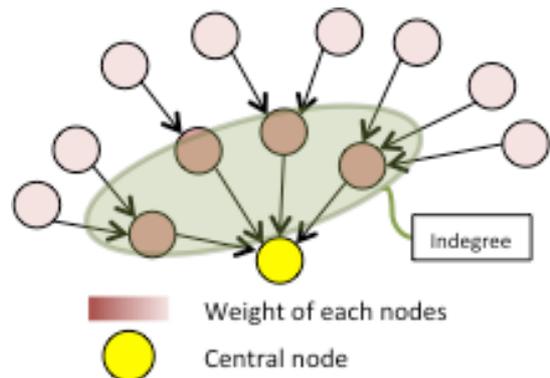


図 7 Harmonic centrality による重み付けのイメージ

#### 3-2. Page views (PV)

英語版 Wikipedia 記事の閲覧回数データを集計し、それぞれの記事への年間アクセス数を集計した結果。なお、Wikipedia Ranking では 2014 年のデータを利用している。

#### 4. 大学に関する記事を対象とした解析

Wikipedia の各言語版では、その言語を母語として用いている者により主要な閲覧・編集がされていると考えられる。したがって、日本語版、英語版 Wikipedia を比較した際、どのような記事で「重要度」が高くなるかは異なる可能性が高い。Wikipedia Ranking を利用して、“university” というジャンルに関する(英語版)記事の重要度ランキングを調べた。Harmonic Centrality(HC), Page views(2014 年一年間の結果。PV)の 2 指標による 2015 年 9 月 25 日時点での 30 位までのランキング、および 13 大学の順位および名称を表 2、表 3 にそれぞれ示す。表中の大学名は、記事のタイトルと一致している。

Wikipedia Ranking は、日本語版記事に関するランキング表示には対応していない。そこで、Python により独自でプログラムを作成し、日本語版 Wikipedia の大学に関する記事についても Page views と Indegree の 2 指標による重要度を測定した。測定には、Wikipedia が無償で公開している、Wikipedia に関する種々のデータ<sup>7)</sup>を利用した。

Page views の測定にあたり、まず対象となる記事を選定するため、前述のデータのうち、日本語版全記事(2015 年 9 月 8 日 16 時 33 分時点)のタイトルが記述されたテキストファイルから、「大学」という単語を含む記事タイトル 6000 件以上抽出した。この内、教育機関としての大学とは直接関係ないタイトルをスクリプトおよび手動で除外し、5256 件に絞った。これらの記事について、Wikipedia の閲覧回数を月別で集計し json 形式で提供しているサイト<sup>8)</sup>を利用し、2014 年の閲覧数を記事毎に集計した。

Indegree の測定にあたり、Wikipedia Ranking による Page views の結果(表 2 右欄)における海外大学上位 25 校に関する日本語版記事、独自で測定した Page views の結果(表 4 右欄)における【タイプ A: トップ型】を除く国内大学 12 校、および【タイプ A: トップ型】の 13 校、の合計 50 大学を対象とした。Indegree 算出の方法としては、ある記事に前述の 50 大学に関するそれぞれの記事へのリンクが存在する場合、各大学のスコアを加算していく。リンクの解析に使用したデータは日本語版 Wikipedia 全記事の最新版(2015 年 9 月 3 日 4 時 50 分時点)、1,972,775 件を記録したものである。この内、一定以上のコンテンツをもつ記事のみを考えることとし、10,000 byte 以上の容量をもつ記事 172,887 件(全体の 8.7%)を抽出し、これらの記事を対象としてリンクの解析と Indegree の算出を行った。

日本語版 Wikipedia に関する Page views(2014 年)、Indegree によるランキング、および【タイプ A: トップ型】13 校の順位と指標の値を表 4、表 5 にそれぞれ示す。

表 2 30 位までのランキング(英語版 Wikipedia)<sup>5)</sup>

HC	PV (2014)
Harvard University	1 University of California, Berkeley
Columbia University	2 Pennsylvania State University
University of Oxford	3 Harvard University
University of Cambridge	4 Georgia Institute of Technology
Yale University	5 Swarthmore College
Stanford University	6 University of East Anglia
Massachusetts Institute of Technology	7 University of Leeds
Princeton University	8 University of Eastern Finland
University of Virginia	9 University of Oxford
University of California, Berkeley	10 Stanford University
University of California, Los Angeles	11 Massachusetts Institute of Technology
New York University	12 Yale University
University of Chicago	13 University of Cambridge
Cornell University	14 Columbia University
University of Colorado Boulder	15 Cornell University
Eton College	16 University of California, Los Angeles
University of Texas at Austin	17 Princeton University
University of Michigan	18 University of Pennsylvania
University of Pennsylvania	19 Brown University
University of Toronto	20 University of Michigan
University of Paris	21 Duke University
University of Southern California	22 University of Southern California
London School of Economics	23 Dartmouth College
Duke University	24 Texas A&M University
University of Illinois at Urbana-Champaign	25 Eton College
Georgetown University	26 New York University
Johns Hopkins University	27 University of Texas at Austin
McGill University	28 University of Notre Dame
University of Edinburgh	29 University of Chicago
University of California	30 California Institute of Technology

表 3 13 大学に関する記事名と順位(英語版 Wikipedia)<sup>5)</sup>

	HC_Rank	PV_Rank
Hokkaido University	1789	1479
Tohoku University	1459	1155
University of Tsukuba	2119	1202
University of Tokyo	49	158
Tokyo Medical and Dental University	-	-
Tokyo Institute of Technology	810	1225
Nagoya University	1953	1421
Kyoto University	154	578
Osaka University	1157	944
Hiroshima University	2040	2713
Kyushu University	1764	1580
Keio University	360	456
Waseda University	189	315

表 4 30 位までのランキング(日本語版 Wikipedia)

Indegree		PV (2014)
東京大学	1	東京大学
早稲田大学	2	早稲田大学
慶應義塾大学	3	慶應義塾大学
京都大学	4	日本大学
日本大学	5	国際基督教大学
明治大学	6	明治大学
法政大学	7	ハーバード大学
ハーバード大学	8	京都大学
筑波大学	9	立命館大学
立命館大学	10	東海大学
東北大学	11	防衛大学校
大阪大学	12	法政大学
北海道大学	13	同志社大学
東海大学	14	筑波大学
オックスフォード大学	15	関西学院大学
九州大学	16	青山学院大学
ケンブリッジ大学	17	近畿大学
名古屋大学	18	大阪大学
同志社大学	19	上智大学
青山学院大学	20	中央大学
東京工業大学	21	東京工業大学
上智大学	22	東京芸術大学
マサチューセッツ工科大学	23	関西大学
コロンビア大学	24	東洋大学
スタンフォード大学	25	一橋大学
広島大学	26	首都大学東京
近畿大学	27	九州大学
イェール大学	28	成蹊大学
関西学院大学	29	立教大学
カリフォルニア大学バークレー校	30	東京農業大学

\*1 橙色は【タイプA:トップ型】13校、青字は海外の大学  
 \*2 PVIは「英語版Page views上位25の海外大学に関する日本語版記事」「日本語版Page views上位12の国内大学に「【タイプA:トップ型】13校」の計50記事を対象としている

表 5 13 大学に関する記事名と順位(日本語版 Wikipedia)

	I_Rank	Indegree	PV_Rank	PV (2014)
東京大学	1	3745	1	339819
早稲田大学	2	2790	2	271416
慶應義塾大学	3	1839	3	249078
京都大学	4	1730	8	172056
筑波大学	9	991	14	142613
東北大学	11	871	38	106075
大阪大学	12	804	18	129433
北海道大学	13	781	41	101887
九州大学	16	651	27	121575
名古屋大学	18	642	42	100222
東京工業大学	21	617	21	126401
広島大学	26	475	66	75250
東京医科歯科大学	41	174	80	68037

## 5. 結果と考察

英語版 Wikipedia に関する結果では Harmonic Centrality, Page views いずれの重要度の指標でも日本の大学はランキングの上位に入っておらず、University of Tokyo(東京大学)が HC 値では 48 位、PV 値では 159 位とそれぞれ最高で、それ以外では 1000 番台、2000 番台になっている大学が大半を占めることが読み取れる。一方、日本語版 Wikipedia に関する結果では、Indegree に関しては海外大学も見られるものの、2 指標ともに国内大学に関する記事がランキング上位の大半を占めていることが読み取れる。

## 6. まとめと今後の課題

当稿では、記事中のリンクの数、つまり編集者の態度に依存する指標である HC, Indegree を利用して「重要度」を測定した。編集、閲覧の双方において、日本語版、英語版ではそれぞれの語圏に属する大学に関する情報がより重視されていると言える。

当稿で取り上げた解析は、サイズが 10,000 byte を超える Wikipedia 記事のみを対象としている。また、重要度に関しても Indegree の算出のみに留まっているため、Wikipedia Ranking による結果と単純に比較できるかといった点には検討の余地が残る。そこで、解析する範囲の拡大の検討や、Harmonic Centrality の値の算出を今後の課題として挙げる。

### 【参考文献】

- 1) 文部科学省. 平成 26 年度「スーパーグローバル大学創成支援」採択構想の決定について  
[www.mext.go.jp/b\\_menu/houdou/26/09/1352218.htm](http://www.mext.go.jp/b_menu/houdou/26/09/1352218.htm)
- 2) 首相官邸. 成長戦略第 2 弾スピーチ.  
[www.kantei.go.jp/jp/96\\_abe/statement/2013/0517speech.html](http://www.kantei.go.jp/jp/96_abe/statement/2013/0517speech.html)
- 3) Wikipedia:全言語版の統計  
[ja.wikipedia.org/wiki/Wikipedia:全言語版の統計](http://ja.wikipedia.org/wiki/Wikipedia:全言語版の統計)
- 4) Kotaro Nakayama, Takahiro Hara, Shojiro Nishio. 2007. Wikipedia Mining for An Association Web Thesaurus Construction, In In Proceedings of IEEE International Conference on Web Information Systems Engineering, 322-334 (2007)
- 5) The\_Open\_Wikipedia\_Ranking:  
<http://wikirank.di.unimi.it/index.html>
- 6) Yannick Rochat. 2009. Closeness Centrality Extended To Unconnected Graphs. The Harmonic Centrality Index, Applications of Social Network Analysis 2009
- 7) Index of /jawiki/latest/.  
<https://dumps.wikimedia.org/jawiki/latest/>
- 8) Wikipedia article traffic statistics. <http://stats.grok.se>

\*1 筑波大学大学院システム情報工学研究科 博士前期課程  
 \*2 筑波大学システム情報系 教授