

ビッグデータ処理技術を用いた Wikipedia マイニング

プロジェクトマネジメントコース・ソフトウェア開発管理グループ 矢吹研究室 1242005 石井康之

1. 序論

Wikipedia は、多くの人がボランティアで執筆するオンライン百科事典プロジェクトである。

Wikipedia は 2001 年 1 月 15 日に創設され、2001 年 5 月ごろに日本語版が発足した。

多くの人が参加するプロジェクトの代表例である Wikipedia を調査することによって、このような形式のプロジェクトのマネジメントについての有意義な知見が得られることが期待できる。

このオープンなオンライン百科事典プロジェクトの成功理由について様々な考察がされており、その中の 1 つの要因に「適切な時に、それぞれのニッチに対しボトムアップとトップダウンの適切な混合率を出していた」というのがある [1][2]。Wikipedia では、ボトムアップの力、無統制の善良さで成長したと見えるのだが、実際にプロセスをよく調べてみると、見かけ以上に入念なトップダウンのしくみによる管理がなされている。

Wikipedia の編集者の中には、管理者と呼ばれる利用者たちがおり、この利用者たちが統制を行っているのではないかと考えた。

そこで当研究では、管理者の動向を見るため、管理者の編集回数がどのように変化しているか調査する。結果から Wikipedia の成功理由に、この管理者がどのように関係しているか見つけ出す。

2. 目的

Wikipedia を一つのプロジェクトとみなし、このオンライン百科事典で管理者の動向がどのように変化しているか調査する。

3. 手法

以下のとおり手法を行う。

1. Wikipedia 日本語版の編集履歴まで含んだファイルをダウンロードし、ローカルでデータマイニングを行う。
2. Wikipedia の管理者の編集回数の変化を解析する。

3. 管理者の編集の割合がプロジェクトの動きにどのようにつながっているのか調査する。

4. 結果

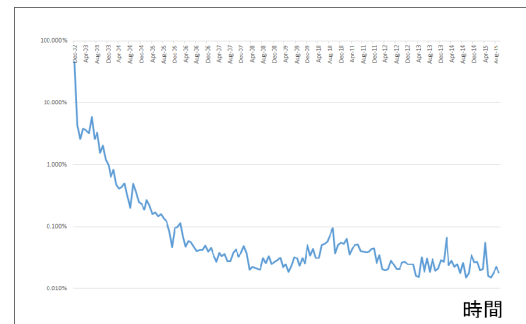


図 1 1ヶ月ごとの総編集回数における管理者の編集の割合の変化

日本語版 Wikipedia が実装されてからの、管理者の月別編集回数のデータ解析を行った。1ヶ月ごとの総編集回数における管理者の編集の割合を出した。

5. 考察

日本語版 Wikipedia では、管理者が編集を行う必要性が薄れてきている。実装当初から、管理者は毎月約 300～400 回ほどの編集を行っていたが、2015 年 5 月ごろからは、月に 100 回も行わないことが多くなってきているからである。

6. 結論

日本語版 Wikipedia では、管理者が編集を行う割合は年々減少している傾向だった。

参考文献

- [1] ケヴィン・ケリー(著). The bottom is not enough, 2008. <http://kk.org/thetechnium/the-bottom-is-n/>.
- [2] 堺屋七左衛門堺屋七左衛門(訳). 七左衛門のメモ帳. <http://memo7.sblo.jp/article/16282081.html> (2015.10.03 閲覧)。