

Wikipedia によるオープンなプロジェクトの品質管理のあり方について

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. 研究の背景

Wikipedia は多くのボランティアにより、始まってから 10 年足らずの間に、大きな成長を見せたオンライン百科事典プロジェクトである。Wikipedia の規模は膨大だ。文字列は 10 億語を超え、ブリタニカとエンカルタの合計の何倍にもなる。英語版の Wikipedia は 250 万以上の記事があるが、全体の 3 分の 1 にも満たない。Wikipedia はさまざまな言語が参加するグローバルなプロジェクトでもある。[1] このオープンなプロジェクトの百科事典は制限なく、誰でも自由に使用でき編集することもできる。だが、編集者のすべてが善意を持っているとは限らず、中には悪意のある編集をするものもいる。悪意のある行為をする人とわかっていても Wikipedia では規制などをしたりはしない。それにもかかわらず、我々が Wikipedia を使用している際はそのような記事は見かけず、信頼のおける品質が保たれている。記事は完成・確定されることはないので、新しい情報にいつでも改変することができる。Wikipedia は信頼されるオープンなプロジェクトの百科事典、それも質量ともに史上最大の百科事典を、顔や素性もわからない人たちと共同作業で創り出されるプロジェクトで、その成果である百科事典本体である。本研究では、Wikipedia の全編集データをマイニングすることによって、Wikipedia の品質が保たれている成功理由を見つけ出す。

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品質管理がどのように行われているか調査する。この調査により、オープンな共同作業プロジェクトにおける、品質管理マネジメントの在り方についての知見を得たい。

3. 研究方法

Wikipedia 日本語版の編集履歴まで含んだ超巨大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い、どのような品質管理が行われて

いるか調査する。また、オープンなプロジェクトにおける品質管理マネジメントの在り方を提案する。

4. 成果物のイメージ

Wikipedia で行われている差し戻しの編集回数を調査しその傾向をグラフにまとめる。さらに、オープンなプロジェクトでの品質マネジメントの知見を得る。

5. 進捗状況

Google が提供している BigQuery というビッグデータを扱うことができるサイトで、Wikipedia のデータを提供していているので、差し戻しデータを抽出し、ランキング化することができた。BigQuery が提供しているデータは、英語版だけのものであり、多言語版を解析するには別の解析方法をする必要がある。[2]

6. 今後の計画

Wikipedia が提供している Wikimedia というサイトから日本語版の全履歴データをダウンロードする。ローカルで解析するためにパソコンの環境も整える必要がある。Wikipedia の全履歴データを解析し、オープンなプロジェクトをする際の品質管理のあり方について調査し提案する。また、既に世に出ている Wikipedia マイニングについて調査を行う。[3]

参考文献

- [1] アンドリュー・リュー. ウィキペディア・レボリューション 世界最大の百科事典はいかにして生まれたか. 株式会社早川書房, 2009.
- [2] Bigquery. <https://cloud.google.com/bigquery/?hl=ja> (2015.09.03 閲覧).
- [3] Wikipedia マイニング. <http://sigwp.org/ja/index.php/Wikipedia%E3%83%9E%E3%82%A4%E3%83%8B%E3%83%B3%E3%82%B0> (2015.09.10 閲覧).