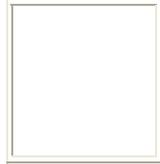


個人が発信する評価情報ページ検索システムの提案

大谷 紀子 研究室

指導教員
承認印

0131162 西潟一生



1. 研究背景と目的

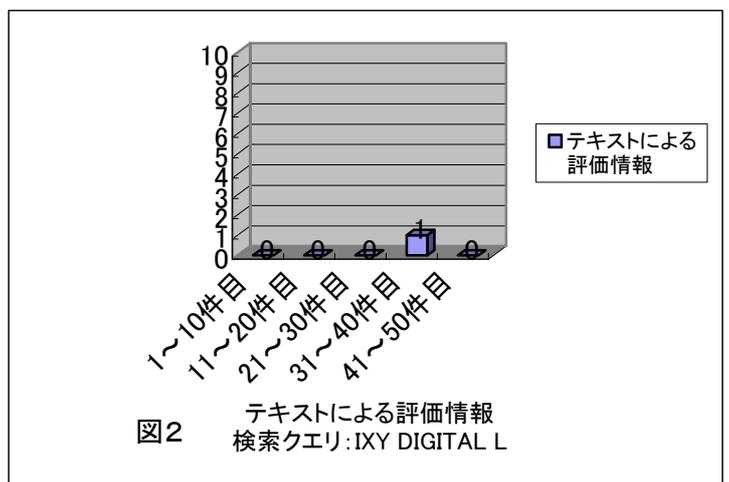
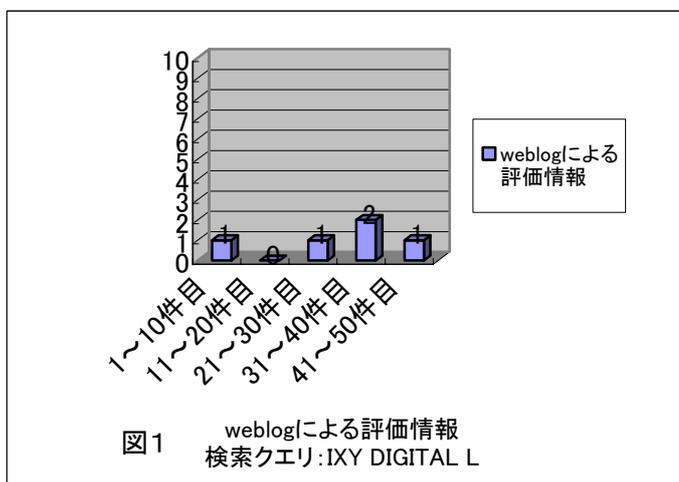
我々が製品を購入しようとした際に、その製品に関する様々な評価情報を考慮し、複数の候補から一つの製品に絞っていくことは誰もが経験することである。価格ドットコムや Amazon に代表されるレビューページの存在とその利用度を見れば明らかであるように、昨今では、製品に関する様々な評価情報の収集をインターネットに頼ることが非常に多い。個人が発信する情報は、対象製品のデメリッスを臆せず載せてある分、メーカーなどの団体が作成したウェブページよりも有意義であると考えられるからである。

本研究ではインターネット上の評価情報検索を目的とし、個人が発信する評価情報の判定方法を提案する。ならびに、その提案手法により、個人が発信する評価情報と判定されたページを対象とする検索システムを構築し、その有効性を検証する。

2. 評価情報ページの判定方法

個人が発信する評価情報収集の判定方法として、以下の3つが有効な判定方法として考えられる。

- (1) : 個人のページであることを表す、URL にチルダ(~)が含まれるウェブページを取得する方法。
- (2) : Google の検索結果の中の上位 50 件の約 10%が、Weblog による評価情報ページであることから、Weblog を取得する方法。(図 1 参照)
- (3) : Google の検索結果の中の上位 50 件の約 0.5%が、テキストのみで構成されている評価情報ページであることから、テキストのみで構成されているページを取得する方法。(図 2 参照)



3. 評価ページ検索システムの構築

Weblog、テキストのみのページの検索については、Google のデータベースを利用し、Google から返された結果をフィルタリングし、目的のページが得られるようにした。また、受け渡しの言語は PHP とした。手順は以下のとおりである。

- 1 : PHP からクエリを生成し、クエリを Google のデータベースに送る。
- 2 : クエリは GoogleAPI を介して Google のデータベースを参照し、結果を返す。
- 3 : Google から受け取った結果を PHP でフィルタリングし表示する。

4. システムの動作結果と考察

図 3 は検索結果を表示させた動作画面例である。

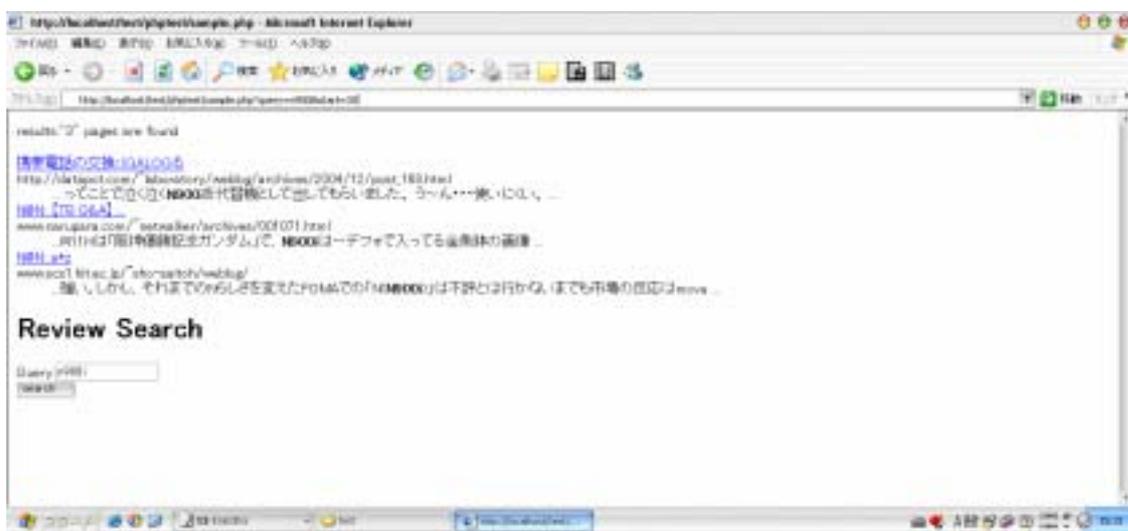


図 3 動作画面例

動作結果より、表示される結果の評価情報ページの割合が増えたことから、評価情報ページを検索する判定方法としては一定の成果を見せたといえる。しかし、本研究の現段階での問題点は多い。

第 1 に、Google から受け取った検索結果の URL を直接開いてソースを読みに行くため、すべての URL を開いてページの判定をするまでに非常に時間がかかり、PHP の関数の制約上、サーバーの処理速度、通信状況によっては処理を正常に実行できず、エラーが発生する場合もある。第 2 に、GoogleAPIs の制約として、検索結果の受け渡しに 1 日 1000 回までしか行えないため、仮に 1 万件ヒットしたとしても、1 万件全てからクラスタリングすることができない。

検索精度の問題として、検索クエリに対する全ての Weblog や、テキストのみのページが評価ページとは限らないため、検索クエリによっては全く精度が上がらないと言える。

以上のことから、今後の課題として考えられるのは、検索時間短縮のために、検索ワードが入力されたのと同時に、その検索ワードを監視するスクリプトを同時生成し、独自のデータベースに格納するシステムを作成する必要がある。また、Weblog の収集の精度を上げるため、RSS フィードを使い、インターネット上に存在する Weblog の分析、収集を行う必要があるといえる。

5. 参考文献

[1] Tara Calishain, Rael Dornfest 著, 山名 早人 監訳, 田中 裕子 訳

『Google Hacks ープロが使うテクニック&ツール 100 選』, 株式会社オライリージャパン, 2003.