

イベント監視による活動履歴の自動作成

奥村 哲也[†] 大西 浩太郎[†] 小山 雅史[‡]

[†]奈良工業高等専門学校専攻科 電子情報工学専攻

[‡]奈良工業高等専門学校 情報工学科

1. 研究背景

近年パソコンの普及により、パソコンを使って作業を行う機会・時間が増加している。そのため、いつ・どの様な作業を行っていたかを後々になって把握することは困難になってきた。特にインターネットの発展はめざましく、Web ページから得られる情報量は膨大である。そこで重要になってくるのが、過去に行っていた作業の要点をまとめて利用者に提示する自動要約技術である。

本研究では、利用者の過去の作業内容を要約し、「活動履歴」を自動で作成するシステムの開発を目的とする。

2. 処理の概要

処理の概要を図1と2に示す。文章で記録を残すことができれば、履歴の内容を容易に把握することができる。日本語を母国語とする利用者は、文章を書く際にも日本語で文章を書くという点に注目し、入力した日本語文章を活動履歴作成に利用する。そこで、図1を用いて文章入力を監視し、取得した文章の要約を行う。

Step1. 日本語文章入力の監視

キーボードから入力された日本語文章が、各アプリケーションに送信されるのを監視し、文字列の取得を行う。

Step2. 重要文抽出による内容要約

重要文とは、全文章の中から利用者の活動内容をよりの確に表している文章のことをいう。本

研究では、ある単語の文章内への出現頻度を単語の重要度と定義する。下記に示す文要度の計算を行った結果、重要度が高いとされた文章を重要文として抽出する。

Step3. アプリケーション名・時刻の取得

Step2 で要約された文章がいつ・どのアプリケーションに対し入力されたのかを取得する。

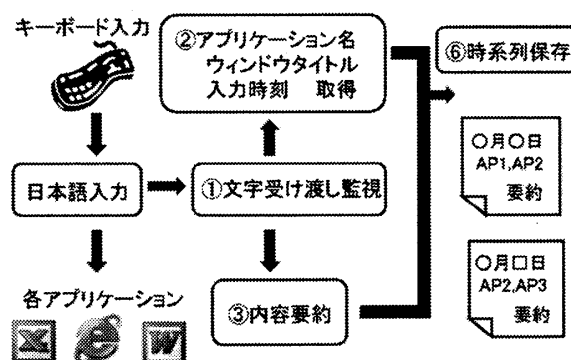


図1 文章入力の監視

入力先のアプリケーションが Web ブラウザである時は、図2を利用し、Web ブラウザの監視を行う。Web ブラウジングは文章入力だけでなく、Web ページに書かれた内容を読むという動作も含まれているため、文章入力に注目するだけでは十分と言えない。また前述の通り現在インターネットには非常に多くの情報が存在するため、利用者の活動を把握するためにはより多くの要素が必要となる。

Step4. 閲覧ページの監視

利用者が Web ページを閲覧した際にアクセスしたページと行動を監視する。

Step5. 利用者固有の重みづけ

アクセスページログから、閲覧したページの HTML タグ・単語出現頻度等を用いた単語重要度の計算を行う。それに対して行動ログから抽出したページ閲覧時間や利用者の注目点情報による利用者固有の重み付けを行う。

Behavior Recording by Event Observation

Tetsuya Okumura[†] Koutaro Onishi[†] Masafumi Koyama[‡]

[†]Advanced Electronic and Information Engineering Course, Faculty of Advanced Engineering, Nara National College of Technology

[‡]Information Engineering, Nara National College of Technology

Step6. 時系列保存

各作業を繰り返し行い、一定時間ごとに活動履歴として出力を行う。

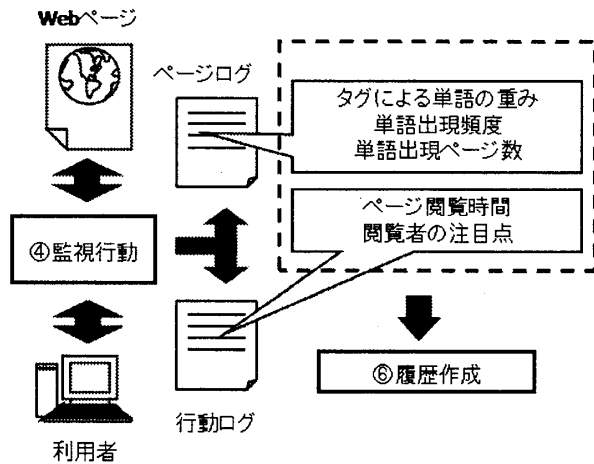


図2 Webブラウザの監視

2.1 単語重要度の計算

キーボードからの日本語文章入力、常に変化するアクティブな情報であり、利用者の活動内容と常に一致している。しかし、Web ページはスタティックな情報であり、利用者の興味により閲覧時間が変化する。これら利用者側に発生する差を利用し、利用者に適した活動履歴の作成を行う。本研究では、アクティブウィンドウの遷移から正確な閲覧時間の計測を行う。その他にも、重要文を抽出し内容要約された活動履歴と、実際に利用者が行った活動との誤差を少なくする方法としては、テキスト中の複数の手がかりを組み合わせたというものがあ[1]、これに準じ単語重要度を以下のように表す。

$$\text{単語重要度} = (\text{time} + \text{save}) \cdot (\text{TF} + \text{html} + \text{input}) \cdot \text{IDF} \dots (1)$$

2.2 文重要度の計算

一文中に含まれる単語の総数を N 、単語重要度を $w_t (t=1 \dots n)$ 、とすると、文重要度は式(2)で表される。

$$\text{文重要度} = \frac{1}{N} \left(\sum_{t=1}^N w_t \right) \dots (2)$$

3. 研究結果と今後の課題

IME (Input Method Editor)が日本語入力の処理を行い、各アプリケーションにメッセージを送信していることを予備実験により確認した。予備実験の結果をもとに、入力された文章が変換・決定された際に発生するメッセージを監視し、IME から各アプリケーションに送られる文字列の取得を行った。取得した文字列をアプリケーション名や入力時刻と共に保存し、一定時間ごとに要約を行った。その結果、図3に示すような日付ごとの活動履歴を自動的に作成することができた。

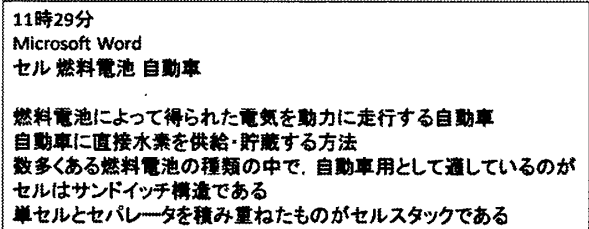


図3 作成された活動履歴

次に、2つの分野 A, B から各 5 ページ合計 10 ページを用意し、被験者にはどちらか片方の分野に対して注目して閲覧させ、その結果得られた動的な情報から、抽出された重要文がどのように変化するかを観察した。表1の結果より、被験者が注目した分野が上位に抽出されることが確認された。

表1 重要文の分野

順位	1	2	3	4	5	6	7	8	9	10
被1	A	A	A	A	B	A	B	B	A	A
被2	B	B	B	B	B	A	A	B	B	B

現在両システムそれぞれの重み付け調整を試行錯誤的に行い、統合化された活動履歴作成を目指している。また、マインドマップ形式等を用い、より分かりやすい活動履歴の提示方法についても検討を行っている。

参考文献

[1]橋本和希“Web履歴を利用した抄録作成に関する研究”平成17年度電気関係学会関西支部連合大会論文G12-5