

Title	創造的活動における文献調査のためのドキュメントスキミング支援環境
Author(s)	羽山, 徹彩; 金井, 貴; 國藤, 進
Citation	人工知能学会論文誌, 19(2): 113-125
Issue Date	2004-01-27
Type	Journal Article
Text version	publisher
URL	http://hdl.handle.net/10119/7843
Rights	Copyright (C) 2004 人工知能学会. 羽山 徹彩, 金井 貴, 國藤 進, 人工知能学会論文誌, 19(2), 2004, 113-125.
Description	

創造的活動における文献調査のための ドキュメントスキミング支援環境

創造的研究活動の支援環境を目指して

Document Skimming Support Environment for Surveying Documents in Creative Activities

Towards Supporting Environment for Creative Research Activities

羽山 徹彩
Tessai Hayama

北陸先端科学技術大学院大学 知識科学研究科
Graduate School of Knowledge Science, Japan Advanced Institute of Science and Technology
t-hayama@jaist.ac.jp, <http://www.jaist.ac.jp/~t-hayama/>

金井 貴
Takashi Kanai

(同上)
kanai@jaist.ac.jp

國藤 進
Susumu Kunifuji

(同上)
kuni@jaist.ac.jp, <http://www.jaist.ac.jp/ks/general/faculty/S.Kunifuji.html>

keywords: document skimming, visualizing interface, support system, recommendation system, personalized summary

Summary

This paper proposes a document skimming environment for surveying documents in our research. Although there are a lot of on-line documents on our surroundings, people, generally, prefer printing out on-line research papers from computer screen. For this reason, although skimming is used for reading documents in our daily life, it is difficult for us to skim documents from computer screens. Therefore, we developed a document skimming environment. The environment has a skimming support system and a recommendation system. The skimming support system supports skimming documents from computer screens by the interface, which is applied the Fisheye effect and the Overview+detail effect. Focus points of the Fisheye effect are the sentences selected by the original sentence extraction algorithm based on the value of standard distribution, and the Overview interface is displayed automatically the generated table of contents. The recommendation system generates personalized summaries by the collaborate filtering, which use users' log of the skimming support system. Furthermore, evaluation results show as follows; The value of F-measure of our sentence extraction algorithm is higher than it of the sentence extraction algorithm based on TF or Japanese lexical chaining method, the skimming support system is more effective method to skim documents from computer screen than paper, and the skimming support environment is more effective method to product research proposal documents than paper.

1. はじめに

多くの創造的活動においては、情報を収集し、分析することでアイデアの創出を行う。ドキュメントはそのようなプロセスにおいて最も重要な情報媒体の一つである。近年の情報ネットワークの発展により、デジタルライブラリーやウェブページから電子化された論文ドキュメントを容易に入手することが可能となったが、我々は通常印刷した論文等のドキュメントを読んでいる。このような傾向の原因としては、ディスプレイからの読み難さの問題がある。近年の研究によると、人間の読むスタイルがはじめから最後まで普通に読む方法から重要どころだけ読むをスキミングへ変化しており [Levy 97]、認知

実験により紙面と比較してディスプレイからのスキミングが苦手であるということがわかってきた [Muter 91]。ここでドキュメントスキミングとは、短時間で正確に文書の意図を理解する行為のことである。

そのため、創造的活動を行うために必要不可欠である文献調査を効率よくするためには、ドキュメントスキミングを支援する環境が必要となる。

ディスプレイから読み易さを支援する研究分野としては、主に視覚効果を利用したインタフェース技術やテキスト要約技術などがある。視覚効果を利用したインタフェースには、主に Overview+detail 効果, Fisheye 効果, Linear 効果がある。Overview+detail 効果は、全体の一覧表示を付加することで、現在の表示部分の位置を知ること

可能にする。Fisheye 効果は、注目する部分（フォーカス）を大きく表示し、それ以外の部分（コンテキスト）を圧縮して表示することで、注目する部分を即座に知ることを可能にする。Linear 効果は、文を順序どおり並べていく表示である。これら視覚効果のディスプレイからの読解に対する特徴として、Overview+detail 効果は理解を支援するが他の効果より多くの時間を要する方法であり、Fisheye 効果はクリティカルタスクに有効であることが知られている [Hornbeck 01]。これら複数の視覚効果を組み合わせてディスプレイからのドキュメントの読み易さを支援する研究として、Suh[Suh 02] と Graham[Graham 99] の研究があるが、いずれのシステムも全体の縮小画像を用いた Overview+detail 効果と Linear 効果を組み合わせてである。

テキスト要約技術は、ドキュメントの量を減らすことで読む負荷を軽減するため、読解を支援する技術であるといえる。しかし、テキスト要約分野では、自動生成した要約の首尾一貫性の問題が指摘されている [Boguraev 98, 奥村 99]。対話的テキスト要約の研究の多くは、重要文抽出手法により提示された要約をもとにしたアプローチを行っている。例えば、TXTRACTOR[McDonald 02] は、単語の頻度、手がかり語、固有名詞と文の位置のパラメータをユーザが調節することでユーザの嗜好に応じた要約を提供する。また、Saggion[Saggion 98] のシステムは、ユーザが論文のアブストラクトに対し読みたい箇所を選択することでユーザの嗜好を反映したアブストラクトを提供する。しかし、これらの研究は、読み手が嗜好に合った要約のために自ら作成する必要がある、作成するための手間が掛かるという問題がある。

そこで本研究では、創造的研究活動における文献調査を支援するために、新たにドキュメントスキミング支援環境を提案する。提案するドキュメントスキミング環境は、効率的な読解を支援する Overview+detail 効果と Fisheye 効果を組み合わせたインタフェースをもつ。

以降では、2 章で本研究のドキュメントスキミング支援環境へのアプローチについて述べる。3 章ではドキュメントスキミング支援環境を構成するシステムの構築について述べ、4 章ではその構築したシステムの評価実験について述べる。最後に 5 章でまとめと今後の課題について述べる。

2. ドキュメントスキミング支援環境へのアプローチ

紙面と比較したディスプレイから読み難さの原因は、一般的にディスプレイが持つ物理的制約の問題と紙が持つ物理的特性の欠如であるといわれている [Mills 87]。

ディスプレイが持つ物理的制約の問題とはディスプレイが紙に比べて視野角を制限することであり、論文など紙に印刷したドキュメントに対してディスプレイ上では一覽

性を確保し難い。そのため、ディスプレイ上においてはスクロールバーを用いた表示方法を行うが、スクロールバーを使用した読みは読み返しが困難であり、また、ページングを使用した読みと比べ内容の理解が難しい [Schwarz 83]。

紙が持つ物理的特性とは、ページ番号や章番号などのドキュメント上に記述されている情報以外に、現在の読み位置を紙の枚数を見ることや触ることでディスプレイに表示するより多くの情報を得ることが可能にする。そのため、ディスプレイで読む場合では読み手が全体に対する読み位置を意識的にドキュメントの記述内容から知る必要があるのに対し、紙からの読む場合ではディスプレイに比べ全体に対する読み位置を把握し易い。

以上の読み難さの原因を克服するために、本研究のドキュメントスキミング支援環境では、以下のようなアプローチを採用した。

- ディスプレイからの読み返しの難しさに対するアプローチ
 - 重要な文に対して Fisheye 効果を適用する
 - ユーザの主観的な判断により各文に対し Fisheye 効果のフォーカス/コンテキストの切替えを可能にする
- スクロールバーを用いた読み難さに対するアプローチ
 - 論文ドキュメントをセグメントごとに表示する
 - 論文ドキュメントの要約を提供する
- ディスプレイからの全体に対する読み位置を把握するためのアプローチ
 - Overview+detail 効果により現在の読み位置の把握を支援する

読み返しが困難な問題に対しては、Fisheye 効果により読み返し箇所を印象付けることを行う。一般に、論文等の読解を行う際には文書に下線を引くことが多いが、この読む過程において下線を引くという行為は読み手が重要であろうと感じた後に読み返すために行う行為である [Marshall 97]。そのため、読み返しをディスプレイ上において支援するアプローチとしては、Fisheye 効果のフォーカスによって予めシステムが重要文に対し印象付け、読む過程において読み手が重要であろうと感じる文に対して対話的に Fisheye 効果の切り替えを可能にする。

次に、スクロールバーを用いた表示方法を行う問題に対するアプローチとしては、ページングと同等の読み易さを実現するために表示するドキュメント量を減らして表示することを行う。表示するドキュメント量を減らすためには、ドキュメントを分割して一度に表示する量を減らすか、ドキュメントの量自体を減らす方法がある。前者においてディスプレイ上からドキュメントを読むために有効なドキュメントの分割単位は、セグメント単位であることが知られている [Pynte 80]。紙面上からの閲覧においても、話題がページによって分割する場合では読

み難く前のページの話題の初めから読み返したり、本や論文において章が変わる場合は改ページを行ったりすることからもセグメント単位の分割が有効であるといえる。そのため、我々は、ドキュメントを分割して一度に表示する量を減らすために、ドキュメントをセグメント単位の分割し表示を行う。また、ドキュメントの量自体を減らす方法としては、ドキュメントの概要を提供する要約がある。要約はドキュメントの内容の詳細まで理解できないが、読み手にドキュメント内容を判断する指標を与える。そのため、ドキュメントスキミング支援環境において要約は、スキミング対象の判断のための補助機能として提供する。

最後に、ディスプレイからの全体に対する読み位置を把握するためのアプローチとしては、Overview+detail 効果をスキミング支援のためにインタフェースへ適用する。Overview+detail 効果は、全体表示を利用することで現在の読み位置を把握する視覚効果である。その全体を表示する Overview には、実世界で用いている目次をメタファーとしたインタフェースを用いる。それによって、我々は、ユーザが全体の概要から現在の読み位置を直感的に理解できることを可能にする。

以上のアプローチをもとに、本研究では、ドキュメントスキミング支援環境を構築する。提案するドキュメントスキミング支援環境は、スキミング支援システムと要約提供システムからなる。スキミング支援システムは、読み返しを支援する Fisheye 効果と現在の読み位置の把握を支援する Overview+detail 効果を組み合わせたインタフェースを提供する。また、要約提供システムは、ドキュメントスキミングのための補助機能として、論文ドキュメントの概要を提供する。システムの詳細については、スキミング支援システムを 3.1 節、要約提供システムを 3.2 節で述べる。

3. ドキュメントスキミング支援環境の構築

3.1 スキミング支援システムの実装

本研究で提案するスキミング支援システムは、Fish-eye 効果と Overview+detail 効果を組み合わせたインタフェースを持つ。Overview+detail 機能においては、Overview に対し目次を自動生成した目次型インタフェースを適用し (図 1 の①参照)、detail に対し Overview の各目次項目 (章、節など) に対応するドキュメントをセグメント単位での表示を行う (図 1 の②参照)。また、本研究では、効率の良い読み返しのために Fisheye 効果を用いるが、予め重要文抽出手法によって抽出した文に対してフォーカスが適用され、ユーザが読む過程において重要と判断した文に対して Fisheye 効果のフォーカス/コンテキストの切替えをマウス操作によって可能にした。

本章では、ドキュメントをセグメント単位の分割するセグメンテーション手法については 3.1 節の §1 で述べ、

Fisheye 効果に用いる重要文抽出手法については 3.1 節の §2 で述べ、Overview に用いる目次型インタフェースについては 3.1 節の §3 で述べる。

§1 セグメンテーション手法

本研究では、セグメンテーション手法として Text Tiling アルゴリズム [Hearst 97] に基づく手法を用いた。Text-Tiling アルゴリズムは統計的手法を用いた最も有効なセグメンテーション手法の一つであり、手がかり語をもとにした手法などに比べ文書内容に依存し難いという性質を持つ。しかし、類似度計算においてベクトル空間モデルを用いているため急激な話題の変化しか検出できなく、論文のような論理的な文脈をもつ文書において話題を検出し難いという問題がある。そこで、本研究のセグメンテーションでは、類似度計算において Jaccard 係数を用いることで論文のセグメンテーションを可能にした McDonald の手法 [McDonald 02] を用いる。本研究では、セグメンテーションを章あるいは節ごとに適用する。これは、章や節が著者の考えによるセグメントであるため、章や節を越えたセグメントがないと考えられるからである。

McDonald の手法は、以下の手順で行う。

1 トークン分割

ドキュメントに対し形態素解析を行い、ストップワードの除去と形容詞、形容動詞、動詞を原型へ変換する。

2 結束スコアの決定

2.1 トークンシーケンスを L づつ区切り、ブロックを作る。

2.2 隣接するブロックの類似度を計算する。

前にあるブロックはそのブロックより前の K トークンシーケンスを加え、後ろにあるブロックはそのブロックより後の K トークンシーケンスを加え、Jaccard 係数 (式 (1)) によりブロック間の類似度を計算する。

$$S_{i,j} = \frac{\sum_{k=1}^L w_{ik} \cdot w_{jk}}{\sum_{k=1}^L w_{ik}^2 + \sum_{k=1}^L w_{jk}^2 - \sum_{k=1}^L w_{ik} \cdot w_{jk}} \quad (1)$$

$S_{i,j}$ はブロック i と j の類似度である。 $\sum_{k=1}^L w_{ik} \cdot w_{jk}$ は、各ブロック i, j に含まれるトークンの頻度の積をすべて加えたものであり、共通して含む数度を表している。

3 境界の決定

求められた結束スコアをトークンシーケンスナンバー、類似度においてプロットし、値の補間を行い滑らかにする。その結果、極小解となった点を境界とする。

本研究のセグメント分割手法に用いたパラメータは、Hearst と同様に L の値を 20、 K の値を 10 とし、ストッ

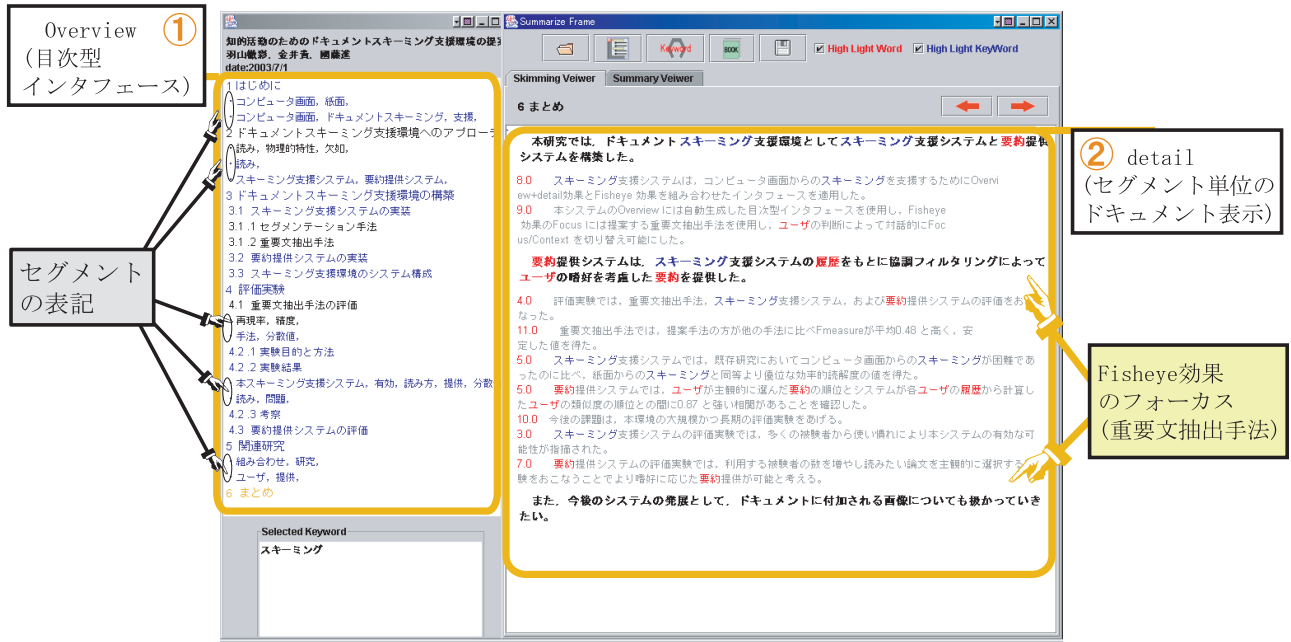


図 1 スキミング支援システムのインタフェース

ブワードは助詞、助動詞と Hearst が用いたワードを日本語化したものを用いた。境界の判定には極小解に近いパラグラフの境目を境界とした。このパラグラフの単位は、空白を空けて文が始まる箇所の前を境としてその境界間に含まれる文書とし、テキスト形式のドキュメントデータからスキミング支援システム用のデータフォーマットへ変換する時にパラグラフを検知する(3.2節参照)。また、形態素解析には「茶筌」[松本 03]を使用し、値の補間には Newton 補間法を用いた。

実際に、本論文の「1. はじめに」をセグメンテーションした結果を図 2 に示す。

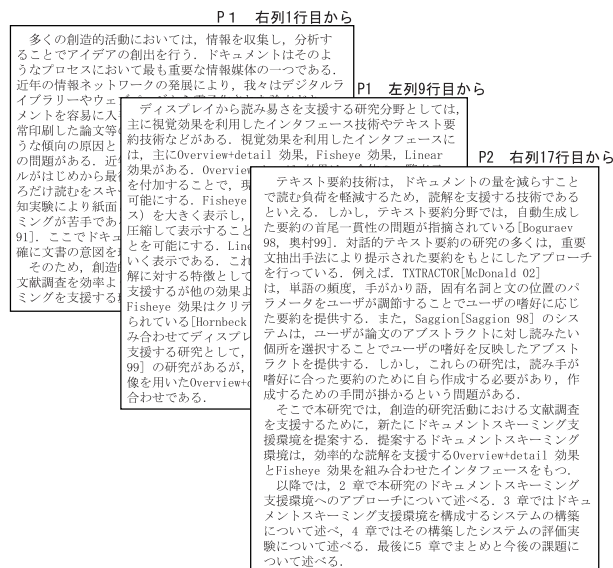


図 2 本論文の1章をセグメンテーションした結果

図 2 が示す結果では、本論文の1章において3つのセ

グメントが検出された。著者が判断するにおいては、本論文の1章のセグメント結果を妥当である。

次に、セグメンテーションによって分割されたドキュメントが、本システムにおいて一覧表示が可能であるかどうかを検証する。論文7本に対しセグメンテーション結果において1セグメントあたりに含まれるワード数(名詞、動詞、形容詞、形容動詞、未知語)を調べたところ、含まれているセグメント数117個においてワード数は約149個含まれていた。その結果より、本システムの一画面に表示できるワード数はフォントサイズ12ptを用いて約450ワードであるため、本手法に分割されたセグメントはほぼスクロールバーを使用しない範囲で論文ドキュメントの表示が可能である。

§ 2 重要文抽出手法

本研究では、ドキュメントのスキミングに焦点を当てているため、重要文抽出において話題の多様性および話題と文章全体の意味関係を考慮する必要がある。ドキュメントスキミング対象においては、『全体に分散する単語は、全体を表現するため重要である。特定の部分にだけ集中する単語は、その部分を表現するため重要である。それら2つのタイプの単語が含まれる文は、全体と部分を結びつけるため重要である』であると仮定し、この仮定に基づいた重要文抽出手法を用いる。用いる重要文抽出手法は、全体を章(あるいは節)とし部分をセグメントとして分散値をもとに文の重要度を求め、手掛かり語を含む文と箇条書きの項目の最初の一文が重要であるというヒューリスティックなルールを加えた。手掛かり語は、人間の判断で選択された論文の重要文(4.1節に用いた正解データ)を元に、一般的に論文の中で著者が主張する上で用いるであろうフレーズである「従って、すなわち、提案する」の3語を用いた。これら3語が人間

の判断で選択された重要文に含まれている割合は、「従って」が7文中6文、「すなわち」が3文中2文、「提案する」が7文中7文と高い割合で選択されている。また、我々は、箇条書きの項目を著者が主張を分類し明確に述べるために使用すると考え、特にその項目の最初の一文が項目の内容を述べていると考えた。実際に、正解データに含まれている割合においても箇条書きの項目の最初の一文は、23文中20文と高い割合で選択されていた。

本論文の2章をヒューリスティックなルールにより重要文抽出した結果を図3に示す。この結果から2章においては、手掛かり語によってP3ページ目の右25行目の文が抽出され、「箇条書きの一文目」によってP2ページ目の左19行目、25行目、28行目の文が抽出された。

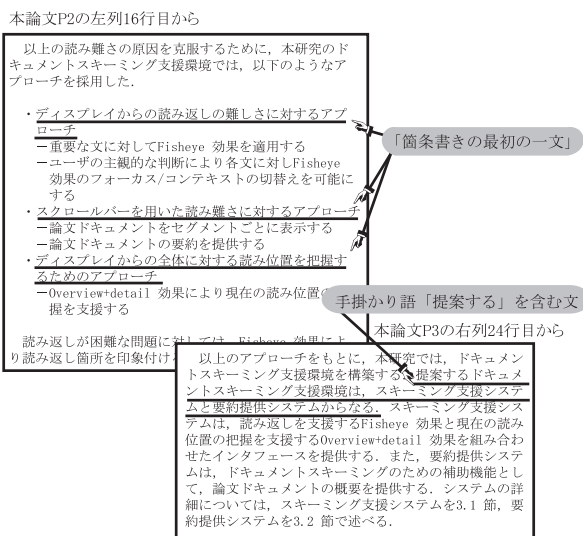


図3 本論文の2章をヒューリスティックなルールにより重要文抽出した結果

以上の本重要文抽出手法の計算手順を以下に示す。

- 1 手掛かり語「従って、すなわち、提案する」を含む文と項目の最初の一文を重要文として抽出する。
- 2 章（あるいは節）とそれに含まれるセグメントとの分散値を使った式(2)によって単語の重みを計算する。

$$\text{単語 } x \text{ の重み} = \left| \sum_{k=0}^N \frac{(x_k - \bar{x})^2}{N} - \sigma^2 \right| \quad (2)$$

k はセグメントナンバー、 x_k はセグメントナンバー k の単語の頻度、 \bar{x} は章（あるいは節）内の単語の総頻度、 N はセグメント数、 σ^2 は全ての単語の分散値の平均である。この式から分散する単語と集中する単語は高い重みとなる。

- 3 各文の重要度を文に含まれる単語の重みの平均値として計算する。
- 4 抽出した文が各セグメント内の総文数の20%となるまで重要度をもとに文を抽出する。

式(2)を用いた重要文抽出手法は、高頻度で分散している単語と低頻度で分散している単語ともに含まれている文だけでなく、高頻度で分散している単語のみが含まれている文あるいは低頻度で分散している単語のみが含まれている文が重要文として抽出される。これら抽出された文は全体の主張を表す文や各セグメントの主張を表す文である。我々は、本手法の重要文の仮説で述べた文以外に全体の主張を表す文や各セグメントの主張を表す文も重要文であると考えた。そのため、式(2)を利用することで、重要文抽出が可能となる。

以上から抽出された重要文は、予めシステムがユーザへ提供する Fisheye 効果のフォーカスとなる。

§3 目次型インタフェースの自動生成手法

目次型インタフェースは、論文ドキュメントデータから章タイトル、節タイトル、あるいは、セグメントタイトルを抽出し、出現した順にアウトライン表記したものである。章タイトルや節タイトルは論文ドキュメントデータにおいて章番号（「1.、2.、」）や節番号（「1.1、1.2、」）を手掛かりに抽出することを行った。セグメントの位置を示すセグメントタイトルはセグメントごとに含まれるキーワードとし、章あるいは節に複数セグメントが存在する場合にはセグメントごとに含まれるキーワードを順に並べたものをそれぞれ表記し、章あるいは節に1つのセグメントしか含まれない場合には目次型インタフェース中において表記しない（図1の①参照）。また、目次型インタフェース上に表記した章タイトル、節タイトル、セグメントに含まれるキーワードは、対応するドキュメントと1対1の対応付けを行い、現在の読み位置を把握しやすいように detail に表示されているドキュメントに対応する目次の項目を色付けした。セグメントに含まれるキーワードは、各セグメント中に出現する名詞において最も高い頻度の単語を含む一文を選択し、その文に含まれる名詞を並べて表示した。これは、人に印象を与えることにおいて語の並びが重要であると考えられるからである。

3.2 要約提供システムの実装

論文の概要を与えることは、その論文の本文を読むかどうかあるいはどの部分を読むべきかの判断のための指標を与える有効な手段であり、その概要が興味や関心に基づく読み手の観点を含むことがより有効な指標を与える[奥村 99]。そのため、要約提供支援システムでは、ユーザの観点に応じた要約を提供することでユーザに読むための指標を与えることを目的とする。要約提供支援システムのアプローチとしては、スキミング支援システムにおいてユーザが読み返しに使用した文をもとに最も観点が近いユーザを特定し、そのユーザが読み返しに使用した文の集合を要約として提供する協調フィルタリングを行う。読み手の観点的の違いによって選択する重要文が異なることから[難波 98]、我々は観点的に近いユーザ同士

が読み返しにおいて選択した文の位置が類似していると考え、最も観点が近いユーザを特定するためにスキミング支援システムにおいてユーザが読み返しに使った文をもとに行う。ユーザが読み返しに使用した文は位置とキーワードを含んでいるため、各ユーザの読み返しに使用した文の類似度によって観点の近いユーザを特定することが可能である。要約提供システムが提供する要約は、最も観点の近いユーザが読み返しに使用した文を集合したものである。そのため、観点が近いユーザは読み返しにおいて同じ文を選択し易く、より近い観点を含んでいる。協調フィルタリングの方法は、以下の手順で行う。

1. 要求するドキュメントを読んだことのあるユーザを検索する。
2. 検索されたユーザ間の類似度を計算する。

ユーザ間の類似度には、式 (3) の Jaccard 係数を用いる。

$$Jaccard(a, b) = \frac{\#focus(a \cap b)}{\#focus(a \cup b)} \quad (3)$$

$Jaccard(a, b)$ は、ある論文のスキミング支援システムの履歴である a と b の類似度を示す。 $\#focus(a \cap b)$, $\#focus(a \cup b)$ は、その履歴 a と b において共通して Fisheye 効果がフォーカスとなっていた文の数あるいは少なくとも一方が Fisheye 効果がフォーカスとなっていた文の数を示す。

3. 類似度の高いユーザを選定し、そのユーザのドキュメントの要約を提供する。

構築した要約提供システムを図 4 に示す。

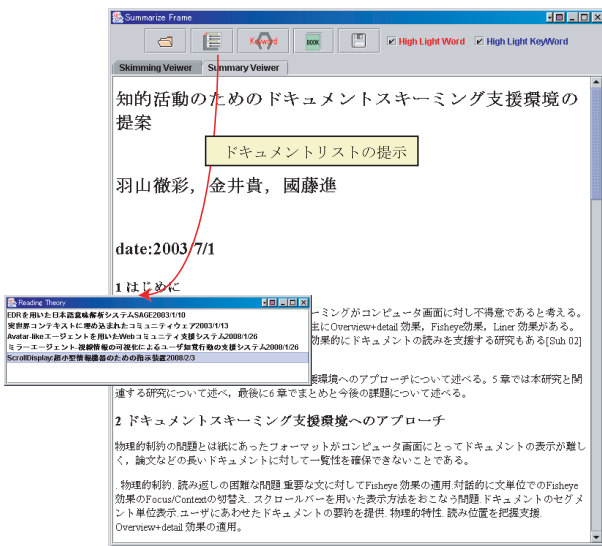


図 4 構築した要約提供システム

要約提供システムが提供する論文ドキュメントの要約は、スキミング支援システムのインタフェースのタブボタンを切り替えることで表示する。ユーザは、サーバー

に登録されている論文ドキュメントリストから論文タイトルを選択し、サーバーにおいて選択されたドキュメントを協調フィルタリングにより類似度計算を行うことで類似度の高いユーザの要約を提供する。

また、ドキュメントスキミング支援環境における要約提供システムの位置づけは、より効果的なドキュメントスキミング支援環境を実現するためにスキミング支援システムの補間的な役割を行うシステムである。提供する論文要約はどの論文あるいは論文の中のどの部分を読むべきかの指標を与え、論文のより深い理解を求める場合はスキミング支援システムを使用することで詳細まで理解することを想定している。

3.3 ドキュメントスキミング支援環境のシステム構成構築したドキュメントスキミング支援環境のシステム構成を図 5 に示す。

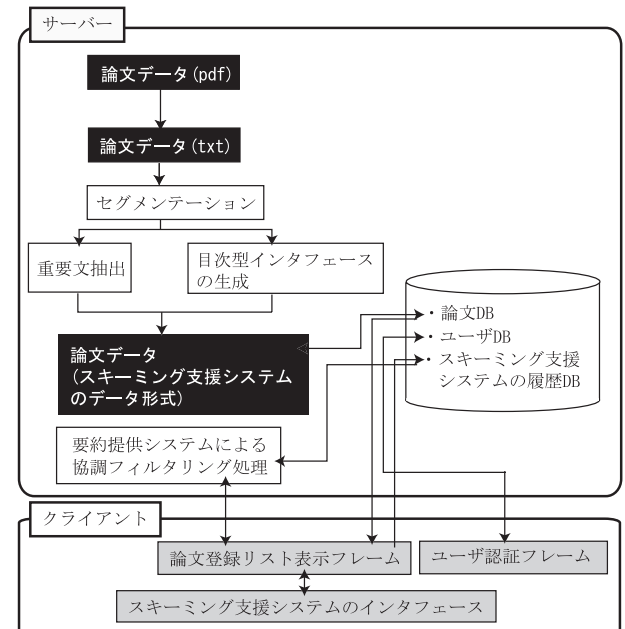


図 5 ドキュメントスキミング支援環境のシステム構成図

構築したシステムは、サーバー側がデータ処理とデータベース管理、クライアント側がインタフェースを提供するサーバー/クライアント・システムである。本環境でのドキュメントファイルの取り込みは Web 上からサーバーへアップロードによって行い、現在は PDF ファイルに対応している。アップロード後、ドキュメントの PDF ファイルはフィルタプログラムによりテキストデータを抽出し、スキミング支援システムに不必要なデータの除去することでテキスト形式のドキュメントデータを生成する。そのドキュメントデータをもとに、セグメンテーション、重要文抽出、目次型インタフェースの生成処理を行い、スキミング支援システムのデータ形式であるドキュメントデータを生成する。これをクライアント側にデータ送信することでスキミング支援システム上に

においてドキュメントの読解を行う。また、テキストデータは文書の記述形式フォーマットで PDF ファイルから抽出されるため、不必要なデータの除去には各文書の記述形式に対応した方法を適用した。本システムでは、現在のところ人工知能学会誌および情報処理学会誌の文書の記述フォーマットに対応している。

データベースには、ドキュメントリスト DB、ユーザ DB、スキミング支援システムの利用履歴 DB が含まれており、ドキュメントリスト DB はアップロードの際に自動登録し、ユーザ DB はユーザを Web 上から登録可能としている。また、スキミング支援システムの利用履歴 DB は、ユーザが使用後に利用履歴、ユーザ名、ドキュメントタイトル名をサーバーへ送信することによって登録される。

クライアント側では、論文ドキュメントのスキミング支援システムのインタフェースを提供する。ユーザはユーザ認証フレームにより、まずユーザ名の入力を行う。それにより、システムはユーザを特定する。次に、ユーザは、論文リストフレームにおいて表示されている論文のリストから、読みたい論文を選択する。その結果、スキミング支援システムのインタフェースでは、図 1 のように要求した論文ドキュメントが表示される。論文リストは、サーバー側の論文 DB に登録されている論文名である。ユーザは、表示された論文ドキュメントを読み進めていく。その際、スキミング支援システムにおける Fisheye 効果は初期段階において重要文抽出により抽出された文をフォーカスしているが、読み進めていく過程においてユーザの判断により文のフォーカス/コンテキストが切り替えられていく。ユーザが論文ドキュメントを読み終えた後、クライアント側のシステムはその論文の最終的な Fisheye 効果のフォーカス/コンテキストの情報とユーザ名をサーバー側へ送信し、その情報はスキミング支援システムのユーザ履歴 DB へ登録される。また、要約提供システムの要約表示では、ユーザ認証後、まずユーザがスキミング支援システムのタブボタンを切り替えて論文リストから読みたい論文の選択し、サーバー側ではそのユーザのスキミング支援システムの利用履歴を用いた協調フィルタリング処理を行う。その結果、クライアント側のインタフェースでは、図 4 のようにユーザに特化した論文ドキュメントの要約を表示される。

4. 評価実験

評価実験では、用いた重要文抽出手法、スキミング支援システム、および、スキミング支援環境について行った。被験者は、すべて北陸先端科学技術大学院大学の学生である。

4.1 重要文抽出手法の評価

本実験の目的は、提案した重要文抽出手法の精度を評価することである。評価方法は、人手で作成した正解データをもとに論文 3 本 (A, B, C) に対し本手法と簡易要約器 Web Posum[望月 02] と比較した。Web Posum は、統計的手法に基づいた複数の基本的な重要文抽出手法を組み合わせることが可能なテキスト要約ツールである。本重要文抽出手法も同様に統計的手法に基づいた重要文抽出手法であるため、評価実験において Web Posum を比較対照として用いることが適切であると考えた。正解データは、評価者 7 人が個別に抽出した重要文に対し過半数 (4 人以上) の人が選択した文とした。実験に用いた 3 つの論文の章の数、節の数、文の総数、正解データにおける文数を表 1 に示す。

表 1 正解データの各論文に含む情報および本重要文抽出と一致した文の数

論文 A	Sec*1	1	2	3		4		5	6					
	Sub*2		1	2	3	4	1	2	3	4				
	SN*3	6	5	7	8	11	11	6	5	6	6	3		
	CS*4	1	2	1	3	1	1	0	1	1	2	4	0	
	MS*5	0	1	1	1	1	1	0	0	1	1	2	0	
論文 B	Sec	1		2	3	4		5		6				
	Sub	1	2			1	2		1	2	3	1	2	
	SN	34	15	16	39	1	8	18	14	9	6	5	4	9
	CS	5	4	8	14	1	4	4	2	3	2	4	4	7
	MS	3	3	2	6	1	1	2	2	2	2	1	1	2
論文 C	Sec	1	2	3		4		5	6					
	Sub		1	2			1	2	3					
	SN	12	14	6	7	13	11	2	11	20	20	16	3	
	CS	5	2	2	2	9	0	0	6	2	3	4	2	
	MS	3	2	1	1	2	0	0	2	1	2	2	1	

*1 章ナンバー、*2 節ナンバー、*3 章あるいは節に含まれる文数、*4 章あるいは節に含まれる正解データの文数、*5 本手法が正解データと適合した文数

表 1 に示すように各論文に含まれる文の総数における正解データの割合は、論文 A において 33 %、論文 B において 21 %、論文 C において 24 % となっている。

重要文抽出手法の精度に関する評価尺度としては、テキスト要約の評価で用いられる式 (4)、式 (5)、式 (6) の再現率、精度、F-measure を用いた。再現率は抽出された文にどれだけ漏れがないかを表す尺度であり、精度は抽出された文にどれだけ無駄な文が含まれていなかを表す尺度である。F-measure は、トレードオフの関係である再現率と精度を考慮した尺度である。

$$\text{再現率} = \frac{\text{適合した文数}}{\text{正解データの文数}} \quad (4)$$

$$\text{精度} = \frac{\text{適合した文数}}{\text{正解データに適合した文数}} \quad (5)$$

$$F - \text{measure} = \frac{2 \times \text{再現率} \times \text{精度}}{\text{再現率} + \text{精度}} \quad (6)$$

また, Posum の適用手法は, 出現頻度を用いた手法と分類語彙表による単語間のつながりを利用した手法の 2 つに対して行い, 要約率は章または節ごとに 20 % とした. 評価結果を表 2 に示す.

表 2 本提案手法と簡易要約器 Posum の再現率, 精度, F-measure の値

論文	システム	再現率	精度	F-measure
A	Posum (TF)	0.25	0.18	0.21
	Posum (TJ)	0.24	0.24	0.24
	本提案手法	0.52	0.45	0.48
B	Posum (TF)	0.44	0.39	0.41
	Posum (TJ)	0.44	0.39	0.41
	本提案手法	0.45	0.50	0.47
C	Posum (TF)	0.15	0.27	0.20
	Posum (TJ)	0.14	0.24	0.18
	本提案手法	0.46	0.56	0.48

表 2 において, Posum(TF) は単語の頻度をもとにした Posum の重要文抽出手法を示し, Posum(TJ) は分類語彙表をもとにした Posum の重要文抽出手法を示す. 表 2 の結果では, 本提案手法が F-measure が平均 0.48 と, 比較した他の手法より有効かつ安定した値であることがわかる.

本重要文抽出手法がスキミングに及ぼす有効な効果としては, 各章や節ごとに有効な指標を提供することである. 表 1 が示すように, 各章あるいは各節における文の数に対する正解データの数是一定の割合でなく, 正解データが含まれない章や節もある. そのため, 本提案手法である精度・再現率ともに平均約 48 % という数字の有効性を示すことは難しい. しかし, 章あるいは節において本提案手法と正解データと一致する文が論文 A の 1 章と 4.1 節以外に 1 つは存在する. このことから, 本提案手法が抽出した文は, それぞれの章や節の理解への指標を与えている可能性が高い.

以上から, 分散値を用いた本提案手法はスキミング支援システムのフォーカスに適用するための有効な手法であるといえる.

また, 分散値のみを用いた手法が有効であるかの評価も行った. 評価方法は, 上記と同様の方法で行い, Posum の頻度を用いた手法と比較した. 分散値のみを用いた手法の評価結果を表 3 に示す.

表 3 分散値のみを用いた手法の評価

論文	再現率	精度	F-measure	Posum(TF) と一致する文の割合
A	0.24	0.27	0.25	0.33
B	0.33	0.33	0.33	0.50
C	0.31	0.49	0.38	0.15

分散値のみを用いた手法の結果は, 表 3 と表 2 を比較すると Posum に比べ 3 本中 2 本が有効な結果, かつ安定した結果を得た. また, 分散値のみを用いた手法と Posum の頻度を用いた手法とそれぞれが抽出した重要文の一致度は, 結果から異なる文を抽出しているといえる.

4.2 スキミング支援システムの評価

§ 1 実験目的と方法

本実験の目的は, スキミング支援システムがディスプレイからのドキュメントスキミングを支援しているかの評価である. 実験は, 普通に読んだ場合とスキミングを行う場合に分けて, 被験者 18 人に対し実施した. 実験環境は, 21 インチ CRT ディスプレイを使用した. 実験方法は, 論文 3 本 (D, E, F) に対し異なる文書閲覧方法で読み, 各論文に対しての 7 問の正誤問題を解くことを行った. 文書閲覧方法は紙, サムネール付き Acrobat Reader, およびスキミング支援システムであり, 各被験者は合計 3 本の論文を読み, それぞれの論文に対し各文書閲覧方法の 1 つを用いた. 論文の割り当て方はランダムに行い, 最終的に論文ごとに各文書閲覧方法のサンプル数が同じであるようにした.

出題した問題は, 論文の主要なフォーマットである「背景・目的」「アプローチ」「評価」「考察」を含むように作成した. ドキュメントを対象とした実験のため, 論文には, 図, 表, 概要を取り除いたものを用いた. 普通に読んだ場合とスキミングを行った場合の異なる点は, 普通に読んだ場合に被験者がかかった平均時間の 3 分の 2 をスキミングを行った場合の制限時間としたことである. この平均時間の 3 分の 2 という値は, 事前実験によって定めた. Muter の既存実験では, 小説に対し普通に読んだ場合の平均時間の 3 分の 1 時間をスキミングを行った場合の制限時間に設定した. しかし, 本研究において対象とするドキュメントを論文に設定したため文書内容の質を考慮し, 事前実験を通して独自の時間設定を行った. 事前実験は, 被験者 8 人に対し論文 2 本を用いて紙面上から普通に読んだ場合とスキミングを行った場合の読解時間を測定した. 被験者は, 4 人ずつの 2 グループに分け, 一方のグループは紙面上において論文 1 を普通に読み, 論文 2 をスキミングを行い, もう一方のグループには各論文に対しその逆の文書閲覧方法を行った. その測定時間の結果を表 4 に示す.

表 4 の結果から, 論文 1 と論文 2 のスキミングを行った場合と普通に読んだ場合にかかった平均時間は異なり, 論文 1, 論文 2 の各タスクにおいて測定された被験者の読解時間の範囲も個人差がある. そのため, スキミングの時間設定を決めることは難しいが, 本実験においては論文 1 と論文 2 の普通に読んだ場合の読解時間に対するスキミングした場合の読解時間の割合の平均が 0.69 であり, その値に近い 0.66, つまり普通に読んだ場合の平均読解時間の 3 分の 2 時間をスキミングした場合の

表 4 事前実験における普通の読解時間とスキミング時間の測定結果

タスク	測定内容	論文 1	論文 2
スキミングを行った場合	平均時間 (秒)	690	716
	測定時間の範囲 (秒)	570-900	650-795
普通に読んだ場合	平均時間 (秒)	1079	966
	測定時間の範囲 (秒)	930-1200	775-1200
普通に読んだ場合の読解時間に対するスキミングした場合の読解時間の割合		0.64	0.74

設定時間と定めた。

被験者に対する実験の事前説明においては、普通に読む場合において被験者の読解方法がスキミングにならないように「最初から一文づつ読む」ことを説明し、スキミングする場合において被験者の時間制限のある普段の読解方法がスキミングであるため設定した制限時間だけを伝えた。

評価方法は評価値を用いた定量的評価とアンケートをもとにした定性的評価によって行った。定量的評価は、各場合において紙面、サムネール付き Acrobat Reader、本スキミング支援システムの3つの方法を評価指数によって比較した。評価指数は、論文を読む時間と理解度を考慮した Jackson の読解に費やした時間割合度 [Jackson 76] をもとにして、各論文の文書と問題の難易度を考慮し各論文ごとに時間の正規化を行った値に対し問題の正解数で割った値を用いた。

本評価実験では、被験者ごとに3つの文書閲覧方法をそれぞれ3本中1本の論文ヘラダムに割り当て、最終的に各文書閲覧方法ごとに3つの論文の読解に費やした時間割合度の平均によって比較を行っている。この方法は、Muter の既存実験と同様であり、文書閲覧方法を対象ドキュメントヘラダムに割振ることで被験者の読解能力を考慮している。しかし、既存実験では同じ小説の異なる箇所を対象ドキュメントとしていたため文書内容の難易度による読解時間を考慮する必要がなかったが、本研究では異なる論文を対象ドキュメントとしているため文書の難易度を考慮する必要がある。そのため、本評価実験では Jackson の評価指数において読解時間を論文ごとに正規化した値を用いた。

また、アンケートを実験終了後に行い、自由回答で記入してもらった。

§ 2 実験結果

普通に読んだ場合とスキミングを行った場合の読解に費やした時間割合度の結果を表 5 に示す。

読解に費やした時間割合度は正解 1 問あたりに費やした読解時間を表現しているため、結果の小さい文書閲覧方法が効率的な読解方法を提供していることを意味する。

表 5 の結果が示すように普通に読んだ場合では本スキミング支援システムが最も有効な文書閲覧方法を提供しているが、分散分析の結果では $0.034 (< F_{0.10}(2, 24))$ と

表 5 普通に読んだ場合とスキミングを行った場合の効率的読解度の結果

実験方法	文書閲覧方法	読解に費やした時間割合度の平均
普通に読んだ場合	P^*	6.71
	$AR + T^{**}$	6.78
	SSS^{***}	6.51
スキミングを行った場合	P^*	6.09
	$AR + T^{**}$	8.00
	SSS^{***}	5.91

*紙, **サムネール付き Acrobat Reader, ***スキミング支援システム

10%優位水準において優位差がなかった。スキミングを行った場合でも本スキミング支援システムが最も有効な文書閲覧方法を提供しているが、分散分析において結果が $2.85 (> F_{0.10}(2, 24))$ と 10%優位水準で優位差を確認した。

表 6 スキミング支援システム使用後のアンケート結果

項目	評価結果
問いの適切さ (5 段階評価)	4
文字の大きさ (5 段階評価)	4.06
文字の読み易さ (5 段階評価)	3.47
システムの使い易さ (5 段階評価)	4.65
システムの応答時間 (5 段階評価)	4.94
有効な文書閲覧方法 P^*	10 人
(18 人中) $AB + T^{**}$	0 人
SSS^{***}	8 人

*紙, **サムネール付き Acrobat Reader, ***スキミング支援システム

スキミング支援システムの定性的評価結果として、実験後のアンケート結果を表 6 に、被験者 18 人の自由回答のアンケート結果を以下に示す。

- 使い慣れればシステムが一番有効だと感じた (8 人)
- 読み返しに有効であった (5 人)
- 目次型インタフェースが直感的で使いやすかった (3 人)
- 色やフォントのカスタマイズをしたい (2 人)
- Fisheye 効果の注目する文が印象に残った (2 人)
- セグメント単位での表示が読みやすかった (1 人)
- セグメント分割が間違っている箇所が読みづかった (1 人)
- フォントが汚い (1 人)

アンケート結果では、スキミング支援システムの提供する機能についての評価を得た。自由回答において、「読み返しに有効なシステム」や「Fisheye 効果の注目する文が印象に残る」といった Fisheye 効果による読み返しのアプローチが効果的に機能していたと考えられる。目次型インタフェースは、「直感的に分かりやすかった」といった意見が得られた。また、Overview として被験者の使用ログからは、Acrobat Reader に対しサムネール

よりページングボタンを使用していたにも関わらず、本システムに対して被験者すべてがページングボタンより目次型インタフェースを使用して読み進めていた。セグメント分割においては、「読み易かった」と「読み難かった」といった賛否両論を得たが、これは Text Tiling 手法の精度の問題であると考えられる。

§3 考察

本実験の普通に読んだ場合において、3つの読み方の読解に費やした時間割合の有意差はなかったが、これは Muter の実験と同様である。スキミングの場合においても Muter の実験同様、有意差を確認した。Muter の実験ではディスプレイより紙面の方が読解に費やした時間割合において 41% 有効であった結果に比べ、本システムではサムネイル付き Acrobat Reader より 35% 有効、紙より 3% 有効である結果であった。本実験により、本スキミング支援システムがディスプレイからのスキミングにおける効率は、紙を用いた場合と同等であると考えられる。

定性的評価結果では、本システムが提供する目次型インタフェースを用いた Overview+detail 効果と重要文に対する Fisheye 効果がドキュメントのスキミングに有効であったといえる。また、実験後の有効な方法として、紙からの文書閲覧方法が 18 人中 10 人と最も高い値を得た。しかし、本システムの習熟度がドキュメントスキミングをより有効にするという 8 人の自由回答結果から、長期使用後の評価実験では本システムが最も有効な方法となる可能性がある。そこで、本システムが使い易くなるまでの時間と労力への調査として新たに「システムを使用頻度に対する読解に費やした時間割合の変化」を調べた。この実験では、被験者 4 人がスキミング支援システムを使って定めた順番どおりに論文 12 本を読み、その論文に関する問題を解くことで、各論文ごとに読解に費やした時間割合を求めた。出題問題は、本評価実験と同様に各論文に関する 7 問の正誤問題とした。また、追加実験における読解に費やした時間割合は、各論文ごとの相対評価をみるため正規化を行っていない。追加実験結果を図 6 に示す。

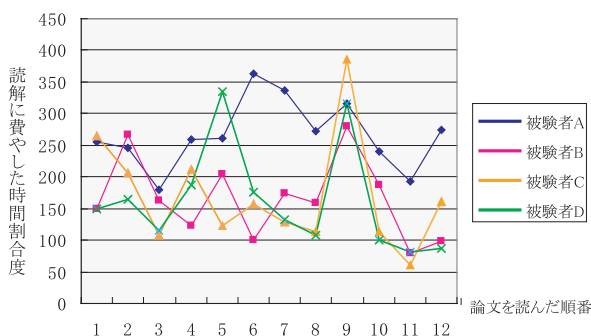


図 6 システムを使用頻度に対する読解に費やした時間割合の変化

各被験者の読解力およびシステムへの慣れる速度が異なるため、読解に費やした時間割合の単純な比較は難しい。しかし、各論文の内容には難易度があるため、システムに慣れた被験者は各論文の内容の難易度に応じて読解に費やした時間割合が変化すると思われる。そのため、すべての被験者が読解に費やした時間割合の変化が論文に応じて増減が同調し始めた時点でシステムの慣れと判断した。図 6 では、論文 7 本目あたりからすべての被験者の読解に費やした時間割合の変化が同調し始めている。以上から、被験者 4 人がスキミング支援システムの習熟のために必要な利用回数は、おおよそ 7 回目以降であると考えられる。

4.3 ドキュメントスキミング支援環境の評価

§1 実験目的

本実験の目的は、構築したドキュメントスキミング支援環境が研究活動において文献調査に対する効果について評価することである。また、本評価実験においてドキュメントスキミング支援環境を実際に使用した結果をもとにスキミング支援システムと要約提供システムの併用方法について考察する。

§2 事前準備

ドキュメントスキミング支援環境に含まれる要約提供システムにおいては、各ユーザのスキミング支援システムの履歴を利用した協調フィルタリングにより他のユーザが作成した履歴を要約として提供する。そのため、本評価実験における事前準備としては、協調フィルタリングに利用する各被験者のスキミング支援システムの履歴と提供する要約のための被験者以外の履歴を用意する。

事前準備として各被験者は、4本の論文をスキミング支援システムを利用して読むことを行った。その際、各被験者は、読解に費やした時間割合を測るために、各論文を読んだ後その論文に関する 7 問の正誤問題を解くことを行った。この読解に費やした時間割合は、本実験におけるグループ分けに利用する。要約作成者は、被験者と異なる 4 人であり、本実験において提供するための 4 本の論文と被験者が事前準備として読んだ論文 4 本の計 8 本をスキミング支援システムを利用して読むことを行った。以上の結果をもとに、本評価実験では各被験者ごとに異なる論文要約を提供する。

§3 実験方法

実験方法としては、被験者 8 人に対し、4 人が紙媒体を、残りの 4 人が本環境を論文資料閲覧方法として利用して制限時間 90 分の中で研究提案書を作成することを行った。研究提案書に対するテーマは「あなたの生活の役に立つウェアラブルについて」とし、被験者へはこちらが用意した「ウェアラブル」に関する 4 本の論文を参考資料として、それ以外の資料を使用しないように指定した。紙媒体を使用したグループと本環境を使用したグループへの被験者の分類方法は、事前実験において求め

た読解に費やした時間割合の結果に基づき、上位から順番に振り分け、最後の4人に対しては各資料閲覧利用者の読解に費やした時間割合の合計が同程度になるように調節した。被験者は、博士前期課程1年の学生6人と博士前期課程2年の学生2人である。本学では、博士前期課程1年の11月まで授業があるため被験者の1年生はまだ研究を始めておらず、被験者の2年生2人はいずれも自分の研究分野がテーマとは異なる。そのため、被験者すべてが「ウェアラブル」分野に対しての既存知識が少ないため、より論文を閲覧すると思った。実験前の被験者に対しては、研究提案書の内容を「背景・目的・期待する効果」を中心に、記述量の目安としてフォントサイズ10.5ptでA4用紙1枚程度とするように説明した。事前実験により、被験者とは異なる学生にこれら4本の論文の読解時間を測定した結果平均約60分という結果を得た。そのため、制限時間90分は、論文の読解60分、文書作成時間30分を想定し設定した。ドキュメントスキミングの評価実験のため、紙媒体の参考資料では表や図を取り除いたものを用いた。また、文書作成には、普段被験者が使い慣れているMicrosoft Wordを指定した。

また、被験者には実験後にアンケートを行い、資料閲覧方法としてシステムを用いた被験者の場合においては各論文ごとのスキミング支援システムと要約提供システムの閲覧時間をシステムの履歴として記録した。

本実験の評価としては、被験者が作成した研究提案書の内容とドキュメントスキミング支援環境の使用方法について行った。研究提案書の内容の評価は、八木下の提案するAHPを使った文章内容の評価する方法[八木下98]によって行った。一般的に、文章内容の評価するには主観的な評価方法に大きく依存するが、この評価方法では比較的客観的な評価が可能である。AHPを用いた文書内容の評価方法は、AHPの一般的な計算手順である階層図の作成、一対比較行列の作成、要素ウェイトの計算、最終目標から見た代替案の総合ウェイトの順に行う。本評価実験では、最終目標を「研究提案書の評価」と設定し、使用した評価項目は「創造性、便利さ、個人的魅力、一般的魅力、具体性、実現可能性、応用可能性」の7項目、整合度の値は0.15とし、評価項目・整合度ともに八木下と同じものを用いた。また、八木下はテーマの異なる文書内容の評価していたため実際の代替案と異なるものを用いていたが、本実験ではテーマを統一して行ったため各研究提案書を代替案として用いた。文書内容の評価には、本学の助手1人、博士後期課程の学生2人、および博士前期課程の学生1人の計4人(A~D)で行った。評価者の博士前期課程の学生は、研究分野がユビキタス分野である。また、ドキュメントスキミング支援環境の使用方法についての評価は、システムの記録とアンケートをもとに考察した。

§4 実験結果

評価者(A~D)が被験者(a~h)によって作成された研究提案書をAHPで評価した結果を表7に、各資料閲覧方法における各評価項目の平均値の結果を表8に、実験後の被験者のアンケート結果を表9に示す。

表9 研究提案書作成の実験におけるアンケート結果(5段階評価の平均値)

アンケート項目	紙	本環境
時間が足りたか	2.75	4.75
資料を使ったか	3.00	3.75
資料に目を通したか	3.50	4.75
資料が有効と感じたか	3.00	4.50
資料の内容を加えたか	3.75	4.50
作成した文書の満足度	2.75	3.50

表7が示すように、本環境を使用した被験者は、評価者の評価結果の平均値において4人中3人が過半数の人より高い値であり、その中の事前に測った読解に費やした時間割合が良い2人は制限時間の3分の2時間で終了している。表8からは、本環境を使用した被験者と紙を使用した被験者とを比べると「具体性」、「可能性」において両者の差がそれほどなかったが、「独創性」、「便利さ」、「一般的魅力」、「個人的魅力」において両者に大きな差があることがあり、本環境を利用した被験者の方が優位な結果が得られた。表9のアンケート結果では、すべての項目において本環境によって資料を閲覧した被験者の方が紙によって資料閲覧した被験者に比べ有効な値であった。特に「時間が足りたか」、「資料に目を通したか」に対する項目においては、紙を利用した被験者に比べ大きな差があった。

次に、本環境を利用した各被験者の閲覧している論文資料とシステム上の表示の時間における推移を表す棒グラフを図7に示す。

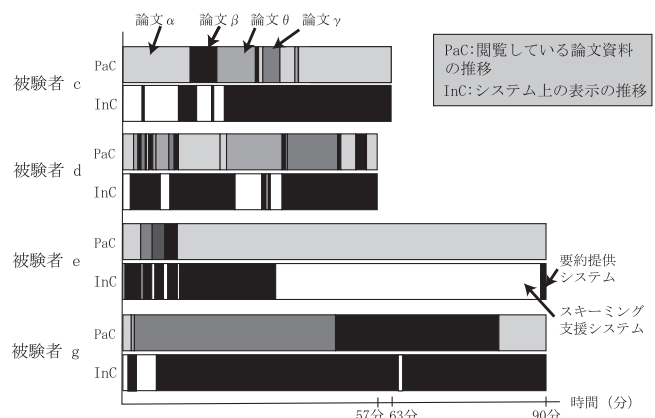


図7 各被験者の閲覧している論文資料とシステム上の表示の時間における推移

図7の各被験者の閲覧している論文資料の推移では、

表 7 AHP による研究提案書の評価結果

作成者	事前実験における 効率的読解度	資料閲覧 方法	作成に掛かっ た時間(分)	評価者 A	評価者 B	評価者 C	評価者 D	評価者の評価 結果の平均値	平均値に基 づく評価順位
a	181	紙	90	0.07	0.14	0.08	0.20	0.07	6
b	171	紙	90	0.10	0.11	0.09	0.10	0.08	5
c	161	本環境	63	0.34	0.08	0.19	0.26	0.15	1
d	129	本環境	57	0.03	0.20	0.14	0.05	0.10	4
e	444	本環境	90	0.07	0.08	0.08	0.06	0.06	8
f	158	紙	90	0.11	0.17	0.11	0.08	0.10	3
g	211	本環境	90	0.24	0.09	0.21	0.14	0.14	2
h	329	紙	90	0.03	0.12	0.10	0.11	0.07	7

表 8 資料閲覧方法における各評価項目の評価結果

	資料閲覧方法	創造性	便利さ	個人的魅力	一般的魅力	具体性	実現可能性	応用可能性
AHP による評価 値の平均値	本環境	0.17	0.16	0.17	0.16	0.14	0.11	0.14
	紙	0.08	0.09	0.09	0.10	0.11	0.15	0.10

すべての被験者が論文資料を頻繁に切り替える時間帯があり、その場合のシステム上においては要約表示を行っていることがわかる。また、システム上にある一定時間の表示を行う論文資料に対し、被験者はスキミング支援システムを使用していた。

§5 考察

本評価実験では、創造的研究活動の文献調査におけるドキュメントスキミング支援環境の評価として、短時間において研究計画書の作成を行い、紙と比較することで評価した。表 8 が示すように、本環境を使用して作成した研究提案書は、紙を使用して作成した研究提案書に比べ「独創性」、「便利さ」、「一般的魅力」、「個人的魅力」において高い内容の評価を得た。また、表 7 と表 9 が示すように、本環境を使用した被験者は、紙を使用した被験者に比べ、短時間で論文資料を読んで研究提案書作成した。そのため、ドキュメントスキミング支援環境は、短時間において質の高いアイデアを創出するのに有効であると考えられる。

また、ドキュメントスキミング支援環境を利用したシステムの履歴においては、各被験者が要約を用いて論文資料を閲覧し、限られた論文に対しスキミング支援システムを使用していたことを確認した。このことから、要約提供システムは、短時間でその論文の本文を読むかの判断を与えるための有効な方法を提供しており、スキミング支援システムの補間的な役割であると考えらる。

5. ま と め

本研究では、ディスプレイからの創造的研究活動における文献調査を支援するために、新たにドキュメントスキミング支援環境を提案し、構築した。構築したドキュメントスキミング環境は、効率的な読解を支援する Overview+detail 効果と Fisheye 効果を組み合わせたイ

ンタフェースをもつ。また、ドキュメントスキミングをより効率的に支援するためのサブシステムとして協調フィルタリングを利用した要約提供システムの構築も行った。

評価実験では、Fisheye に適用した重要文抽出手法、スキミング支援システム、およびドキュメントスキミング支援環境の評価を行った。重要文抽出手法の評価では、分散値に基づく本手法の方が他の手法に比べ F-measure が平均 0.48 と高く、安定した値を得られた。スキミング支援システムの評価では、既存研究においてディスプレイからのスキミングが困難であったのに比べ、紙面からのスキミングとほぼ同等の読解に費やした時間割合の値を得られた。ドキュメントスキミング支援環境の評価では、研究提案書作成という創造的な場において紙を利用するよりも短時間で質の高いアイデアを創出することを確認した。

今後の課題としては、本環境の大規模かつ長期の評価実験をあげる。ドキュメントスキミング支援環境の評価実験では限られた論文をもとに行われたが、今後は実際の研究活動における文献調査に用いて評価することにより、より研究活動に特化したドキュメントスキミング支援環境の構築が可能と考えられる。また、今後のシステムの発展としては、論文ドキュメントに付加される画像についても考慮していきたい。

謝 辞

評価実験に御協力頂いた北陸先端科学技術大学院大学知識科学研究科の皆様には感謝致します。

◇ 参 考 文 献 ◇

- [Boguraev 98] B. Boguraev, C. Kennedy, R. Bellamy, S. Brawer, Y. Y. Wong, and J. Swartz: Dynamic presentation of document content for rapid on-line skimming, In Proceedings of AAAI Symposium on Intelligent Text Sum-

- marization, pp118-128, 1998
- [Graham 99] J. Graham: The Reader's Helper: A Personalized Document Reading Environment, In Proceedings of CHI'99, ACM Press, pp481-488, 1999
- [Hearst 97] M.A. Hearst; Texttiling: Segmenting text into multi-paragraph subtopic passages, Computational Linguistics, Vol.23, No.1, pp33-64, 1997
- [Hornbeck 01] K. Hornbaek, E. Frokjaer; Reading of Electronic Documents: The Usability of Linear, Fisheye, and Overview+detail Interfaces, In Proceedings of CHI'01, pp293-300, 2001
- [Jackson 76] M.D. Jackson and J.L. McClelland: Processing Determinants of Reading Speed. Journal of Experimental Psychology: General 108, pp151-181, 1976
- [Levy 97] D.M. Levy; I read the news today, oh boy: reading and attention in digital libraries, Proceedings of the second ACM international conference on Digital libraries, pp202-211, 1997
- [Marshall 97] C.C. Marshall :Annotation: From Paper Books to Digital Library, In Proceedings of the 2nd ACM international conference on Digital libraries, pp131-140, 1997
- [松本 03] 松本裕治 他: "形態素解析「茶筌」version2.3.1 使用説明書", <http://chasen.aist-nara.ac.jp/chasen/>, 2003
- [McDonald 02] D. McDonald and H. Chen: Using Sentence Selection Heuristics to Rank Text Segments in TXTRACTOR, In Proceedings of the 2nd ACM/IEEE Joint Conference on Digital Libraries, pp25-38, 2002
- [Mills 87] C.B. Mills and L.J. Weldon: Reading text from computer screens, ACM Computing Surveys, Vol.19, pp329-358, 1987
- [望月 02] 望月 源; "テキスト簡易要約器 Posum version1.50.2 マニュアル", JAIST Technical Memorandum, 2002
- [難波 98] 難波英嗣, 奥村学: 観点に基づいた新聞記事の重要文選択に関する心理実験と考察, 言語処理学会第4回年次大会併設ワークショップ「テキスト要約の現状と課題」, pp30-35, 1998
- [奥村 99] 奥村学, 難波英嗣: テキスト自動要約に関する研究動向, 自然言語処理「テキスト要約のための言語処理」特集号, Vol6 No.6, 1999
- [Muter 91] P. Muter and P. Maurutto: Reading and skimming from computer screens and books: The paperless office revisited?, Behaviour and Information Technology, Vol.10, pp257-266, 1991
- [Pynte 80] J. Pynte and G. Noizet: Optimal segmentation for sentences displayed on a video screen, In Proceedings of Visible Language, Vol.2, pp376-385, 1980
- [Saggion 98] H. Saggion and G. Lapalme: The Generation of Abstracts by Selective Analysis, In Proceedings of AAAI Symposium on Intelligent Text Summarization, pp23-25, 1998
- [Schwarz 83] E. Schwarz, I.P. Beldie, and S. Pastoor; A comparison of paging and scrolling for changing screen contents by inexperienced users, Human Factors, Vol.24, pp279-282, 1983
- [Suh 02] B. Suh, A. Woodruff, R. Rosenholtz, and A. Glass: Popout Prism: Adding Perceptual Principles to Overview+Detail Document Interfaces, In Proceedings CHI 2002, pp251-258, 2002
- [Voss 99] A. Voss, K. Nakata and M. Juhnke: Concept Indexing, In Proceedings of GROUP '99: International ACM SIGGROUP Conference on Supporting Group Work, ACM Press, pp1-10, 1999
- [八木下 98] 八木下 和代, 宗森 純, 首藤 勝: 内容と構造を対象としたKJ法B型文章評価方法の提案と適用, 情報処理学会論文誌, Vol.39, No.7, pp2029-2042, 1998

〔担当委員: 角 康之〕

2003年7月1日 受理

著者紹介



羽山 徹彰(学生会員)

2001年同志社大学工学部知識工学科卒業。2003年北陸先端科学技術大学院大学知識科学科博士前期課程修了。2003年4月同大学博士後期課程入学。現在に至る。知識処理システム, 創造性支援システムに興味がある。ACM, 情報処理学会, 電子情報通信学会, 日本創造学会 各会員。



金井 貴(正会員)

1994年立命館大学理工学部情報工学科卒業。1996年北陸先端科学技術大学院大学情報科学研究科博士前期課程修了。1999年北陸先端科学技術大学院大学情報科学研究科博士後期課程修了。博士(情報科学)。同年北陸先端科学技術大学院大学知識科学研究科助手。現在に至る。法的推論, アブダクションの研究に従事。情報処理学会, ACM, IEEE-CS, 日本創造学会 各会員。



國藤 進(正会員)

1974年東京工業大学院理工学研究科修士課程終了。同年, (株)富士通国際情報社会科学研究所入所。1982~86年, ICOT 出向, 1992年より北陸先端科学技術大学院大学情報科学研究科教授, 1998年より知識科学研究科教授, 現在は主として発想支援システム, グループウェア, 知識システムの研究に従事。情報処理学会創立25周年記念論文賞, 人工知能学会1996年度研究奨励賞各受賞。博士(工学)。情報処理学会, 計測自動制御学会, 電子情報通信学会, 日本創造学会など各会員。