The University of Osaka
Institutional Knowledge Archive

Title	子供によるWeb閲覧・Web検索の支援に関する研究
Author(s)	岩田, 麻佑
Citation	大阪大学, 2013, 博士論文
Version Type	VoR
URL	https://hdl.handle.net/11094/27485
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

博甲 16080

子供によるWeb閲覧・Web検索の支援に関する研究

2013年1月

岩田　麻佑

子供によるWeb閲覧・Web検索の支援に関する研究

提出先　　大阪大学大学院情報科学研究科

提出年月　2013年1月

岩田　麻佑

関連発表論文

1. 学会論文誌発表論文

1. Iwata, M., Arase, Y., Hara, T., and Nishio, S.: Web Browser for Children Using Bubble Metaphor, *International Journal of Web Information Systems*, Vol. 6, No. 1, pp. 55–73, 2010.

2. 岩田麻佑, 荒瀬由紀, 原 隆浩, 西尾章治郎: 子供による Web 検索のための検索結果リランク手法, 情報処理学会論文誌, Vol. 52, No. 3, pp. 1055–1068, 2011.

3. Iwata, M., Miyamoto, H., Hara, T., Komaki, D., Shimatani, K., Mashita, T., Kiyokawa, K., Uemukai, T., Hattori, G., Nishio, S., and Takemura, H.: A Content Search System Considering the Activity and Context of a Mobile User, *International Journal of Personal and Ubiquitous Computing*, 2012.

2. 国際会議発表論文

1. Iwata, M., Arase, Y., Hara, T., and Nishio, S.: Investigation of Children's Characteristics for Web Browsing, in *Proceedings of International Workshop on Web Information and Data Management (WIDM 2009)*, pp. 91–94, 2009.

2. Suzuki, A., Iwata, M., Arase, Y., Hara, T., Xing, X., and Nishio, S.: A User Location Anonymization Method for Location Based Services in a Real Environment, in *Proceedings of the ACM International Conference on Advances in Geographic Information Systems (GIS 2010)*, pp. 398–401, 2010.

3. Iwata, M., Arase, Y., Hara, T., and Nishio, S.: A Children-Oriented Re-ranking Method for Web Search Engines, in *Proceedings of International Conference on Web Information System Engineering (WISE 2010)*, pp. 225–239, 2010.

4. Iwata, M., Hara, T., and Nishio, S.: How Different are Preferences of Children and

Adults on Web pages?, in *Proceedings of International Workshop on Advanced Distributed and Parallel Network Applications (ADPNA 2011)*, pp. 394–399, 2011.

5. Iwata, M., Hara, T., Shimatani, K., Mashita, T., Kiyokawa, K., Nishio, S., and Takemura, H.: A Location-based Content Search System Considering Situations of Mobile Users, in *Proceedings of International Conference on Ambient Systems, Networks and Technologies (ANT 2011)*, pp. 426–433, 2011.

6. Ashikaga, E., Iwata, M., Komaki, D., Hara, T., and Nishio, S: Exploring Map-based Interactions for Co-located Collaborative Work by Multiple Mobile Users, in *Proceedings of International Conference on Advances in Geographic Information Systems (GIS 2011)*, pp. 417–420, 2011.

7. Mashita, T., Komaki, D., Iwata, M., Shimatani, K., Miyamoto, H., Hara, T., Kiyokawa, K., Takemura, H., and Nishio, S: A Content Search System for Mobile Devices based on User Context Recognition, in *Proceedings of International Workshop on Ambient Information Technologies (AMBIT 2012)*, pp. 1–4, 2012.

8. Mashita, T., Komaki, D., Iwata, M., Shimatani, K., Miyamoto, H., Hara, T., Kiyokawa, K., Takemura, H., and Nishio, S: Human Activity Recognition for a Content Search System Considering Situations of Smartphone Users, in *Proceedings of International Workshop on Ambient Information Technologies (AMBIT 2012)*, pp. 1–2, 2012.

9. Iwata, M., Sakai, T., Yamamoto, T., Chen, Y., Liu, Y., Wen, J-R., and Nishio, S: Aspectiles: Tile-based Visualization of Diversified Web Search Results, in *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 85–94, 2012.

10. Kato R., Iwata, M., Hara, T., Suzuki, A., Arase, Y., Xie, X., and Nishio, S: A Dummy-based Anonymization Method based on User Trajectory with Pauses, in *Proceedings of International Conference on Advances in Geographic Information Systems (GIS 2012)*, 2012.

3. 国内研究会等発表論文（査読付）

1. 岩田麻佑, 荒瀬由紀, 原 隆浩, 西尾章治郎: 泡メタファーを用いた子供向け Web 提示手法について, 情報処理学会マルチメディア，分散，協調とモバイルシンポジウム (DICOMO 2008) 論文集, pp. 553–560, 2008.

2. 岩田麻佑, 小牧大治郎, 荒瀬由紀, 原 隆浩, 西尾章治郎: 泡メタファを用いた子供向け Web ブラウザの評価, 情報処理学会マルチメディア，分散，協調とモバイルシンポジウム (DICOMO 2009) 論文集, pp. 1415–1423, 2009.

3. 岩田麻佑, 荒瀬由紀, 原 隆浩, 西尾章治郎: Web ページの構成と文章に着目した Web 検索結果の子供向けリランク手法, 情報処理学会マルチメディア, 分散, 協調とモバイル (DICOMO 2010) シンポジウム論文集, pp. 1365-1372, 2010.

4. 鈴木晃祥, 岩田麻佑, 荒瀬由紀, 原 隆浩, 西尾章治郎: 位置情報サービスにおける実環境を考慮したユーザ位置曖昧化, 情報処理学会マルチメディア, 分散, 協調とモバイル (DICOMO 2010) シンポジウム論文集, pp. 1434-1441, 2010.

5. 足利えりか, 岩田麻佑, 小牧大治郎, 原 隆浩, 上向俊晃 , 西尾章治郎: 複数の携帯端末を用いた協調作業における地図操作に関する一考察, 情報処理学会マルチメディア, 分散, 協調とモバイル (DICOMO 2011) シンポジウム論文集, pp. 587–594, 2011.

6. 岩田麻佑, 原 隆浩, 嶋谷健太郎, 間下以大, 清川 清, 西尾章治郎, 竹村治雄: モバイルユーザの状況を考慮した位置依存コンテンツ検索支援システム, 情報処理学会マルチメディア, 分散, 協調とモバイル (DICOMO 2011) シンポジウム論文集, pp. 1537–1544, 2011.

7. 嶋谷健太郎, 間下以大, 宮本大樹, 岩田麻佑, 原 隆浩, 清川 清, 竹村治雄, 西尾章治郎: スマートフォンを用いたコンテンツ検索支援のためのモバイルコンテキスト認識, 情報処理学会マルチメディア, 分散, 協調とモバイル (DICOMO 2011) シンポジウム論文集, pp. 1552–1559, 2011.

8. 鈴木晃祥, 岩田麻佑, 荒瀬由紀, 原 隆浩, Xie Xing, 西尾章治郎: ダミーを用いた位置曖昧化手法の評価, 情報処理学会マルチメディア通信と分散処理ワークショップ (DPSWS 2011), pp. 194–199, 2011.

9. 大澤 純, 岩田麻祐, 小牧大治郎, 原 隆浩, 西尾章治郎: スマートフォンユーザのコンテキストと利用アプリケーションの関連性分析, 情報処理学会マルチメディア, 分散, 協調とモバイル (DICOMO 2012) シンポジウム論文集, pp. 1855–1862, 2012.

10. 加藤 諒, 岩田麻佑, 原 隆浩, 鈴木晃祥, 荒瀬由紀, Xie Xing, 西尾章治郎: 停止を伴うユーザの移動経路を考慮したダミーによる位置曖昧化手法, 情報処理学会マルチメディア通信と分散処理ワークショップ (DPSWS 2012), pp. 210–218, 2012.

11. 岩田麻佑, 原 隆浩, 西尾章治郎: 子供による Web 検索支援のための検索クエリに関する子供向けサブトピックの調査, 情報アクセスシンポジウム 2012, pp. 23–30, 2012.

4. その他の研究会等発表論文

1. 岩田麻佑, 原 隆浩, 西尾章治郎: Web ページに対する子供と大人の嗜好に関する調査, 平成 22 年度情報処理学会関西支部支部大会講演論文集 (CD-ROM), 2010.

2. 山本彩奈, 岩田麻佑, 原 隆浩, 荒瀬由紀, Xie Xing, 西尾章治郎: 車両トレースデータを用いたダミー生成による位置プライバシ保護手法, 情報処理学会研究報告 (マルチメディア通信と分散処理研究会 2012-DPS-153), pp. 1–6, 2012.

以上

内容梗概

インターネット環境の普及により，小学生を始めとする子供がWeb上の情報を閲覧することが一般的となっている．さらに，情報を閲覧するだけでなく，Web検索を用いて，Web上の必要な情報を自ら探す子供も増えている．しかし，難解な文章を苦手とし，画像を好むというような子供の特徴を考慮したWebシステムはほとんど存在しないため，現状では子供がWebを十分快適に利用できる環境が整っているとはいえない．具体的には，多数のWebページが一般向けに作成されており，Web上には，子供にとって理解が難しく，面白みに欠ける情報が大多数を占めている．そのような中で，自分の必要とする情報を探す必要があるが，子供は，検索要求を表す検索キーワードを入力したり，検索結果のランキングから必要な情報を選択することを苦手とする．

そこで本研究では，学校の宿題などでWebを利用すると考えられる6才から12才の小学生を中心とした子供を対象とし，子供にとって興味をひき，分かりやすい形式でWeb上の情報を取得できるように，子供によるWeb閲覧・Web検索の支援を行う機構について議論する．具体的には，まず，難解で面白みに欠けるWebページを子供が興味を持ち，理解できる形式に変換する子供向けWebブラウザの検討を行う．次に，Web上の情報を閲覧するだけでなく，Web検索を用いて情報を探す子供を支援するため，子供にとって重要な情報を検索結果の上位にランクする子供向け検索結果リランク手法について検討する．最後に，検索方法についても支援を行うため，子供にとって重要なサブトピックを優先的に提供する子供向けサブトピックのランキング手法の検討を行う．

本論文は，5章から構成され，その内容は次の通りである．まず，第1章において，序論として研究の背景について述べる．

第2章では，Webページの表示形式を改善することで，Web閲覧の支援を行うため，子供のWeb閲覧能力や嗜好を考慮し，子供にとって興味をひき，分かりやすいWebページ提示を実現する子供向けWebブラウザを提案する．提案ブラウザでは，Webページのコンテンツを泡メタファを用いて表示することで，コンテンツの情報を泡の色やサイズなどに反映し，視覚的な情報として伝える．さらに，コンテンツの詳細内容は絵本形式に変換して提示することで，子供が興味を持ち，理解しやすいコンテンツ表示を実現する．これにより，子供が一般向けWebページを興味を持って閲覧することが可能となる．

第3章では，Web検索結果を改善することで，Web検索の支援を行うため，検索エンジ

ンの検索結果を子供向けにリランクする手法を提案する．提案手法では，ページ中の画像量や文章量などのWebページの構成に関する指標，子供向け表現や難解表現の出現数などのWebページの文章に関する指標に基づいてWebページの子供向け度合いをスコア化し，リランクを行う．これにより，子供向けのWebページほど検索結果の上位にランクされるため，子供はランキング下位まで探す手間なしに，子供にとって読みたい，分かりやすいWebページを容易に得ることができる．

第4章では，Web検索方法を改善することで，Web検索の支援を行うため，検索クエリに関する子供にとって重要なサブトピックの調査を行い，子供向けサブトピックランキング手法を提案する．サブトピックとは，検索クエリを具体化するキーワードで，クエリ“沖縄”では“歴史”，“食物”などがサブトピックとして考えられる．子供向けサブトピックの特定に必要な要素を明らかにするため，Wikipediaや検索エンジンなどの複数のソースから抽出したサブトピックを子供にスコア付けしてもらい，調査を行う．そして，調査結果に基づき，取得ソースの種類や難易度を元に，サブトピックを子供向け度合いの降順に並べるサブトピックランキング手法について検討する．これにより，子供にとって調べたい，分かりやすいサブトピックを容易に得ることができる．

第5章では，本論文の成果を要約したのち，今後の研究課題について述べ，本論文のまとめとする．

目 次

1 序章 **1**

1.1 研究の背景 1

1.2 関連研究 2

1.2.1 Web 閲覧・Web 検索に関連する子供の特徴についての研究 2

1.2.2 子供向け Web システムに関する研究 6

1.3 本研究のアプローチ 9

1.3.1 本研究の位置づけ 9

1.3.2 泡メタファを用いた子供向け Web ブラウザ 13

1.3.3 Web 検索結果の子供向けリランク手法 13

1.3.4 Web 検索クエリに関する子供向けサブトピックランキング手法 . . 14

1.4 本論文の構成 15

2 泡メタファを用いた子供向け Web ブラウザ **17**

2.1 まえがき 17

2.2 Bubble Browser の設計と実装 18

2.2.1 設計方針 18

2.2.2 基本動作 21

2.2.3 機能の設計 22

2.3 評価実験 28

2.3.1 実験環境 28

2.3.2 実験手順 30

2.3.3 実験結果 31

2.3.4 考察 35

2.4 むすび 36

3 Web 検索結果の子供向けリランク手法 37

3.1 まえがき 37

3.2 子供向けリランク手法 38

3.2.1 子供向けページの定義 38

3.2.2 指標の設計 39

3.2.3 リランク手順 46

3.3 評価実験 46

3.3.1 データセット 47

3.3.2 評価指標 48

3.3.3 実験結果 48

3.3.4 考察 56

3.4 むすび 60

4 Web 検索クエリに関する子供向けサブトピックランキング手法 63

4.1 はじめに 63

4.2 関連研究 66

4.3 子供向けサブトピックの調査 66

4.3.1 サブトピックのスコア付け 66

4.3.2 調査結果 70

4.4 子供向けサブトピックランキング手法 84

4.4.1 手法の設計 84

4.4.2 評価 86

4.5 むすび 89

5 結論 91

5.1 本論文のまとめ 91

5.2 今後の課題 93

謝辞 95

第1章

序章

1.1 研究の背景

インターネットの爆発的な普及により，インターネットを利用することでWeb上の大量の情報の中から必要な情報を簡単に得ることが可能となった．今や一家に一台PCを所有することが一般的となっており，子供の周りでもインターネット環境が整ってきている．これにより，子供によるインターネットの利用が広まり，Web閲覧・Web検索を行うことが一般的となってきている．2011年の統計[1]によると，小学生の81.9%がパソコンを使用し，そのうち，69.8%の子供が家庭や学校に設置されたパソコンでインターネットを利用している．gooリサーチによる2010年の調査[2]によると，インターネットの利用開始時期は，未就学児で19.2%，小学校低学年で33.8%であり，また，63.0%の小学生が週に30分以上インターネットを利用していることが分かっている．また，別の2010年の統計[3]によると，週に1回以上自宅でパソコンを利用している子供は，小学5年生で53.1%，中学2年生で68.3%である．これらの統計データより，幼少期の頃から，多くの子供がインターネットを用いて日常的にWebを利用していることが分かる．さらに，gooリサーチによる2010年の同調査[2]によると，小学生のインターネットの利用目的は，“ゲーム”が58.9%で最も多く，次いで，“勉強・宿題”に関する検索が52.8%，“趣味・娯楽”に関する検索が51.1%であり，小学生を中心とした子供が，Webページを閲覧するだけでなく，学習や趣味のため

[1]< 内閣府：http://www8.cao.go.jp/>

[2]<NTTレゾナント：http://research.goo.ne.jp/>

[3]< 社団法人日本PTA全国協議会：http://www.nippon-pta.or.jp>

に，Web検索も頻繁に使用していることが分かる．

しかし，難解な文章を苦手とし，画像を好むというような子供の特徴を考慮したWebシステムはほとんど存在しないため，現状では子供がWebを十分快適に利用できる環境が整っているとはいえない．具体的には，多数のWebページが一般向けに作成されており，Web上には，子供にとって理解が難しく，面白みに欠ける情報が大多数を占めている．子供がWeb上の情報を探す際には，そのような子供向けでない情報が大量にある中で，自分の必要とする情報を探す必要があるが，子供は，自分の検索要求を表す検索キーワードを入力したり，検索結果のランキングから必要な情報を選択することを苦手とする．そのため，Web閲覧・Web検索とも，子供にとって，興味をひき，分かりやすく使えるものであるとはいえず，インターネットは子供に学ぶ機会を与えてくれる場であるにも関わらず，十分にインターネットを活用することができないと考えられる [28]．このように，子供のWeb閲覧・Web検索の支援に関する技術の重要性が高まっている．

1.2 関連研究

1.2.1 Web閲覧・Web検索に関連する子供の特徴についての研究

Web閲覧・Web検索に関連する子供の特性として，既存研究で以下のような点が指摘されている．

一般的な情報処理能力

Web閲覧を行う場合，Webページ中の情報を処理し理解する能力が必要である．一般的に，子供は大人と比べると，情報処理能力が劣っている [57]．これは作業記憶容量が小さいためであり，情報を深く理解させるためには，見せたい情報のみに絞り込んで見せるほうが効率がよい [57]．つまり，豊富な散らばった情報の中で全体を見回し，時には目的とは関係のないモノとの偶然の出会いを楽しみながら検索していきたいという欲求（検索性）がある [82]．また，情報処理能力の大きな特徴として，楽しい環境や経験を伴うことで，知識が活性化され学習効果につながることがあげられる [2]．これは特に，幼児教育には必要不可欠な特性である．

また，ポータルサイトにおいてWebサイトがカテゴリごとにまとめられているように，Web上の情報を探す際には，カテゴリという概念が重要である．大人にとってカテゴリは

一般的な概念だが，幼少期の子供はカテゴリ概念，特に抽象的な上位カテゴリの理解を苦手としている [83]．例えば，“犬” の上位カテゴリが “動物” であると理解するのは困難である．このカテゴリ概念の理解をサポートするには，カテゴリの具体的な事例，特徴を示すことが効果的である [87,90]．例えば，“犬” を “馬” などの他の動物の例と同時に示すことで，“犬” が “馬” などの他の動物の例と共有する特徴の把握が促進され，“動物” というカテゴリの理解へとつながる．

Web 閲覧に関する特徴

Web ページに対する子供のユーザビリティに対象を絞った研究が行われている．菊地ら [58,59] は，小学校のパソコンの授業時の子供の様子を観察することで，小学生の Web ブラウジングの特徴を分析している．調査の結果，小さな子供ほど難解な漢字を用いた文章量の多い Web ページを嫌い，キャラクタなどのイラスト中心の Web ページを好むという特徴を明らかにしている．また，小さな子供ほど Web ページ閲覧の途中に飽きる傾向が強いことも指摘している．Nielsen ら [71–74] は，6 才から 12 才，13 才から 17 才の子供に子供向けに作成された Web ページ，大人向けに作成された Web ページの両方を閲覧してもらい，ユーザビリティを調査している．その結果，子供は文字が詰まったようなページよりも見た目に分かりやすいページを好む傾向があることが示されている．特に，6 才から 12 才の子供の特徴として，アニメーションなどのマルチメディア要素やカラフルであるなどのデザイン要素を重視する点，スクロールをほとんど行わず，画面上部に見えている部分のみで操作を行うことがほとんどである点が明らかにされている．しかし，単に見た目が魅力的なだけではなく，コンテンツが多すぎず，できる限りシンプルに子供が操作できるようなデザインが必要であるとも述べられている．さらに，子供が Web ページを閲覧中に他のコンテンツに興味が逸れ，集中して閲覧するのが苦手であることも指摘している．また，富士通による報告[4] では，小学生を対象とする子供向け Web ページを作成する際のユニバーサルデザインについて述べられている．具体的には，子供にとって，分かりやすく，読みやすい文章とするために，文末に語りかけの表現を使用することや，専門用語や難解な漢字，アルファベットを避けること，文章量を少なくすることが必要であり，さらに，イラストなどを用いて見た目に興味を引くように工夫することも重要であると述べられている．

[4]< 富士通キッズ キッズコンテンツ作成ハンドブック: http://jp.fujitsu.com/about/kids/handbook/>

Webページとは異なるが，子供向けの文章として，湯浅 [89] は，子供向けの小学生新聞などの記事と一般向けの新聞の記事を比較し，子供向け文章の特徴を調査している．その結果，子供向け文章は一般向け文章よりも，漢字の割合が少ないこと，1文の文字数が少ないこと，語りかけ表現や話し言葉が多用されていることが明らかになっている．

このように，様々な調査を通じて，子供はカラフルで画像の多いWebページを好み，難解な文章の多いWebページを苦手とするというようなWebページに対する嗜好があることが指摘されている．

Web検索に関する特徴

子供がWeb検索を行う際の特徴を調査した研究も数多くある．Bilalら [9] は，大学院生と7年生（日本の中学1年生）の子供に，ある特定のコンテンツをWeb検索により探すタスクを行ってもらい，大人と子供の間のタスクの成功率や検索行動の違いを調査している．その結果，子供には，検索に失敗したときに次にどうするべきなのか分からない，タスク中でも他に興味のあるコンテンツに気が散ってしまうというような特徴があるため，Web検索によって必要な情報を探すことが大人よりも苦手であることが示されている．また，Bilal [6–8] は，7年生から9年生の子供に，子供向け検索エンジンで，ある特定の答えを探すタスクや自由にコンテンツを探すタスクを行ってもらい，認識面，身体面，感情面での特徴についても調査している．調査の結果，子供は検索するための能力が高くないため，タスクを完了させるのが苦手であり，システムが子供のWeb検索を支援することの必要性が指摘されている．Druinら [16] は，7才から11才の子供にWeb検索によりコンテンツを探すタスク行ってもらう調査により，子供がWeb検索をする際に，スペリング，タイピング，クエリ作成，検索結果の解釈が問題となることを明らかにしている．特に，子供は自然言語でクエリを作成する傾向が強く，検索に失敗しやすいことや，検索結果第1位のページに依存し，上位5件を超えたページを閲覧することがほとんどなかったと示されている．さらに，Druinら [15,27] は，7才，9才，11才のより大規模な人数の子供に自宅で自由な方法でWeb検索を行ってもらい，その様子を観察することで調査を行っている．調査では，年齢，性別ごとに検索を行う際のモチベーションや障害，どのように検索を行うかのプロセスなどについて分析している．その結果，検索意欲はあるが，タイピングの問題で検索に失敗するグループ，普段よくパソコンを使用しており，検索成功率の高いグループ，ある特定のサイトのみで検索を行う傾向のあるグループなどの7つのグループに

子供を分類している．Jochmann-Mannak ら [64] は，近年増えつつある子供向けにデザインされた 4 つの検索インタフェースと一般向けの検索インタフェースである Google を用いて，小学生にタスクを行ってもらい，子供向けにデザインされたインタフェースの効果について検証している．実験の結果，タスクの成功率などについて，子供向けにデザインされたインタフェースと Google の間に差はなく，子供の特徴を考慮したインタフェースが必要であると指摘している．

一方，検索エンジンにおける子供の検索ログの解析を行うことで，子供の Web 検索行動について調査している研究も数多くある．Torres ら [17,18] は，DMOZ の Kids & teens のディレクトリに登録されている Web サイトを訪問している AOL のクエリログを解析することで，子供向け Web サイトを閲覧するユーザの Web 検索行動を調査している．DMOZ とは，世界的ボランティアグループが運営する ODP（Open Directory Project）という包括的ディレクトリサービスであり，人手で整理・分類された Web サイトのデータベースである．また，AOL のクエリログは一般公開されている検索ログのデータである．この調査により，一般向けの Web サイトを目的としているクエリと子供向け Web サイトを目的としているクエリの間にはクエリの長さや数，クエリ発行間隔などに差があることが示されている．Torres ら [19] は，さらに，6 才から 18 才の年齢で Yahoo!のアカウントに登録しているユーザの Yahoo!の検索ログを解析することで，子供の Web 検索行動を調査している．具体的には，自然言語で発行されたクエリの割合，検索に失敗して諦めたセッションの割合，広告をクリックした割合，検索されたトピックの分布，閲覧されたコンテンツの可読性などについて，子供と大人を比較し，子供の方が検索中に混乱し，失敗することが多いことや，ユーザの年齢と閲覧されるコンテンツの可読性に相関があることが示されている．Gossen ら [31] は，ドイツの子供向け検索エンジンのログファイルを解析することで，大人と比較した際の子供の Web 検索行動を調査している．調査の結果，子供が発行するクエリは短く，頻繁にスペルミスをすること，さらに同じクエリでの検索を繰り返し，同じ Web ページへ何度も訪れて，検索中に混乱して失敗する傾向が強いことが明らかになっている．

このように，これまで行われてきた様々な調査により，子供が Web 検索を行う際には，検索クエリの入力や検索結果の選択など多くの問題が存在することが明らかになっている．

1.2.2 子供向けWebシステムに関する研究

子供によるWeb閲覧やWeb検索の支援の重要性が指摘されている中，近年，子供にターゲットを絞ったWebシステムが多く提案されている．

子供向けWeb検索支援システム

子供向けWeb検索システムとしては，PuppyIRという子供向けWeb検索システム作成のためのフレームワークの構築に関する研究 [5,30,60] を始めとし，近年，多くの研究が行われている．まず，子供向けに検索結果のランキングを改善する手法がいくつか提案されている．Thompsonら [88] は，ユーザの年齢に基づいた読解レベルと検索エンジンによって返される検索結果のタイトルと短い文章（スニペット）およびWebページの難易度の差に基づいて，検索結果のランキングを決定する手法を提案している．これにより，子供は自分の読解レベルに応じたWebページを容易に取得できる．Gyllstromら [34] は，あらかじめ定義した子供向けサイトとのリンクの繋がりに基づいてWebページの子供向け度合いを算出し，検索結果のランキングを決定する手法を提案している．これにより，子供向けサイトと少ないステップで繋がっている子供向けである可能性の高いサイトを容易に取得可能となる．

子供向けのコンテンツかどうかをクラスタリングする手法もいくつか提案されている．Eickhoffら [23,24] は，Webページのテキストの複雑さ，読みやすさなどをスコア化し，Webページの集合から子供向けページを特定するための手法を提案している．Polajnarら [77] は，一般的なbag-of-wordに加えて，可読性を評価する手法をいくつか考慮することで，子供にとって適切なニュースコンテンツをどうかを分類する手法を提案している．

さらに，検索結果の表示方法を改善するための研究も行われている．Gyllstromら [32] は，Googleのクエリログ内で，頻繁に共起して出現するキーワードを利用して，クエリに関して適切なメディアタイプ（塗り絵，楽譜など）を決定し，子供向けのマルチメディア検索を実現するシステムを提案している．Glasseyら [25,29] は，子供がよく検索するトピックに関するコンテンツを強調表示し，あまり検索しないトピックに関するコンテンツをフィルタリングするシステムを提案している．このシステムでは，子供が閲覧したコンテンツを記録し，興味のあるトピックについて学習し，個人の興味や特徴に適したコンテンツを優先して提供する．

検索クエリに関する研究も行われている．美馬ら [65,66] は，教科書に基づいたオントロ

ジを利用した検索クエリの拡張方法を提案している．この手法では，小学校の教科書を分析してオントロジーを構築し，オントロジーを元に検索クエリを子供の学習向けに拡張する．例えば，“りんご” を検索クエリに指定した場合，産地である “青森” や，栄養素である “食物繊維” などの関連語により果物としての意味的制約を加える検索クエリを作成する．Nakaoka ら [69] は，幼小者の生活様式オントロジーを構築し，子供の生活環境に密着した検索クエリの拡張を可能とするシステムを提案している．生活様式オントロジーには，幼小者の体験するイベントに関連する事柄を記述する．例えば，“クリスマス” ならば “人気のクリスマスプレゼント” などの具体例を記述する．Polajnar ら [78] は，絵の辞典を用いたクエリにより，Web コンテンツを検索可能にするシステムを提案している．絵の辞典には，基本的な英語の単語が登録されており，自由に変更可能となっている．子供は辞典に定義された絵を選択することでクエリを作成し，クエリに関連する Web コンテンツを検索することができる．Gyllstrom ら [33] は，クエリが意見の分かれる内容であるのか，子供にとって適切なトピックであるのかを特定する手法を提案している．これにより，入力されたクエリが，子供にとって難解なトピックである時はシステムが積極的に支援を行い，意見の分かれる内容のクエリである時は，不適切な検索結果をフィルタリングすることが可能となる．Kammerer ら [56] は，自然言語で入力されたクエリ中の語や句を 1 つのクエリとして扱うことを可能とするクエリ修正のためのツールを提案している．これにより，子供は分かりやすい自然言語のクエリで検索しつつも，クエリ内の語や句を選択することで，それらに関する検索結果を取得することが容易となる．

その他の子供向け Web システム

その他にも子供が Web を利用する際の支援を行う研究が行われている．Sumi ら [85,86] は，子供のコンテンツ内容に対する理解を支援するために，Web コンテンツを絵本に変換するシステムを実現している．電子コンテンツのテキスト情報をもとに，意味的なタグを付加し，オントロジやシソーラスを用いることで，自動的にアニメーションと対話表現によるわかりやすい絵本形式への変換を行う．また，Belder ら [11,12] は，文章を子供にとって分かりやすいように変換するシステムを提案している．このシステムでは，長い文章を分割し，難解な表現は容易な表現に置換することで文章の変換を行う．

Web コンテンツを対象としたものではないが，他にも子供向け Web システムについての研究が行われている．Hutchinson ら [42] は，子供向けの電子図書館のシステム ICDL

(International children's digital library) を提案している．ICDL では，子供が自分の見たい本を簡単に探し，閲覧できるように，子供の能力や嗜好を考慮したインタフェースを提案している．例えば，カテゴリを本のカバーの色やイラストというような子供特有の基準にしたり，カテゴリの階層構造をなくすという工夫をしている．これにより，子供はインターネットを用いて簡単に本を探すことができる．また森ら [68] は，インターネットを通じて世界中の子供たちがコミュニケーションを取れるシステムを提案している．このシステムでは，言葉ではなく，主に子供が実際に頻繁に使用する絵文字を組み合わせて使ってコミュニケーションを取る．さらに，Van ら [84] は，子供向けの健康検索システムを提案している．このシステムでは，キャラクタのイラストと体の部位名をテキストデータとして共に表示し，子供が何をクエリとして入力して検索すればよいのか簡単に思い出せるようにしている．

商用化された子供向け Web サービス

一般に商用化されている子供向けの Web サービスである子供向けブラウザや子供向け検索エンジンがいくつか存在する．子ども向けのブラウザとして既に商用化されているものとしては，以下のようなものがある．KidZui[5] は 3 才から 12 才の子供を対象とした Web ブラウザで，保護者や教師を含む約 200 人以上の編集者が選択した 50 万強の Web サイトのみが閲覧できるようになっている．また，SNS 機能を備えており，保護者の承認を得ると友達を作ることができる．ひらがな・なびぃ[6] は漢字を自動で平仮名に変換するブラウザで，保護者が特定の URL，単語を指定すれば，Web ページのフィルタリングも可能になる．さらに，文章要約機能やアイコンつきのリンク集作成の機能も備えている．Miss America Kid-Safe Web Browser[7] は子供向けにカスタマイズされた教育用 Web ブラウザであり，子供たちが見ても安全と判断された 10,000 以上のサイトにのみアクセス可能である．アニメーションが安全なネットの楽しみ方を言葉でアシストしたり，ポータルを通じて教育用ゲームで遊んだり，仮想マネーを使った様々な活動を行える．Children's Browser[8] は子供がアクセスする Web ページに制限を加えるものである．アクセス可能な Web ページのリストがあらかじめ設定されており，保護者がリストを加減できる機能を備える．

[5] <KidZui: http://www.crunchbase.com/company/kidzui>

[6] < ひらがな・なびぃ: http://jp.fujitsu.com/group/flm/eco/hiranavi/>

[7] <Miss America Kid-Safe Web Browser: http://www.missamericakids.com/>

[8] <Children's Desktop and Browser: http://www.childrensdesktop.com/>

一方，商用化されている子供向けのWeb検索エンジンとして，以下のようなものがある．Yahoo!きっず[9] は，あらかじめ手作業で安全と判断されたサイトのみ検索可能である．一般向けの検索エンジンと同様に，カテゴリやキーワードによる検索が可能である．カテゴリによる検索時には，教科を中心としたカテゴリを利用でき，キーワードによる検索時には，一般向け検索エンジンと同様の関連検索語が最大3語提示される．検索結果のランキングは，サイト検索の結果をまず表示し，その後，ページ検索の結果を表示する．ページ検索結果のランキングは基本的にYahoo! JAPAN[10] と同じであるが，Yahoo!きっずに登録されていないサイトのページはランキングに表示されない．キッズgoo[11] は，Web上のどのようなページでも検索可能であるが，フィルタリング機能を持ち，有害と判断されたページは選択しても閲覧することができない．Yahoo!きっずと同様に，カテゴリやキーワードによる検索が可能となっており，検索結果のランキングは，サイト検索の結果の後にgoo[12] と同じページ検索の結果が表示される．このような子供向け検索エンジンは，有害情報の削除を主な目的としているため，子供は安全にWeb検索を行うことができる．

1.3 本研究のアプローチ

1.3.1 本研究の位置づけ

1.1節，1.2節で述べたように，小学生を始めとする子供がインターネットを利用し，Web閲覧することは一般的となっている．さらに，Web閲覧だけでなく，小学生や中学生では，学校での勉強や娯楽のため，自ら調べたいキーワードを入力し，Web検索を行う子供が増えている[3]．代表的な子供向け検索エンジンであるYahoo!きっずでは，対象を小学生・中学生としており[13]，キッズgooでは，対象を小学生としている[14]．そこで，本研究でも，学校の宿題などでインターネットを使って調べものなどを行うと考えられる6才から12才までの小学生を中心とした年齢の子供を対象とする．また，文献 [15] では，7才，9才，11才の83人の子供によるWeb閲覧・Web検索時の行動の調査の結果，検索意欲はあるが，

[9]<Yahoo!きっず: http://kids.yahoo.co.jp/>
[10]<Yahoo! JAPAN: http://www.yahoo.co.jp/>
[11]< キッズgoo: kids.goo.ne.jp/>
[12]<goo: www.goo.ne.jp/>
[13]<Yahoo きっずガイド : http://guide.kids.yahoo.co.jp/info/index.html>
[14]< キッズ goo ガイド : http://kids.goo.ne.jp/guide/kids_goo/index.html>

Webページの理解や検索クエリの作成の問題で検索に失敗するグループ（Developing），特定のドメインに関するサイトのみを閲覧するグループ（Domain-specific），普段頻繁にパソコンを使用しており，検索成功率の高いグループ（Power），検索するのが嫌いで，何をすればいいのか分からないグループ（Non-motivated），テレビなどの周囲環境によって検索が長続きしないグループ（Distracted），検索中は画像情報ばかりに依存するグループ（Visual），検索結果第1位の閲覧とクエリ修正を繰り返すグループ（Rule-Based）という7つのグループに子供を分類している．これらのグループのうち，Developing グループに，最も多い70%もの子供が属しており，これらの子供のためのWeb閲覧・Web検索の支援の重要性が指摘されている．そのため，本研究では，小学生の子供の中でも，特に，Developing グループに属するような，Webを利用して，趣味や学習に関する情報を閲覧したい気持ちはあるが，自分が見たい情報にうまく辿り着けないような子供の支援を対象とする．

このような子供がWebを利用する際には，Internet Explorer[15] などの一般的なWebブラウザを用いて，Yahoo!きっずやキッズgooの提供する子供向けWebページを始めとし，足りない情報については一般向けWebページを閲覧することで情報を得ると考えられる．自ら検索を行い，必要とする情報を探す際も，Yahoo!きっずやキッズgooといった子供向けWeb検索エンジンを始めとし，足りない情報についてはYahoo! JAPANなどの一般向け検索エンジンを利用して探すものと考えられる．

一般向けに作成されたWebページや検索エンジンは，大人向けに作成されており，子供の特徴を考慮していないため，子供にとって難解で面白みに欠ける情報が多く，子供が興味を持って，それらのサービスを利用することが難しいと考えられる．例えば，子供にとって勉強になるページがあっても，難解な漢字やアルファベットが多用された文章ばかりのページであれば，子供は閲覧する意欲をなくしてしまい，そのページから学習できるはずであった内容を知るせっかくの機会を失ってしまうかもしれない．また，検索エンジンを利用して検索を行う際に，検索の支援を行うために提示される関連検索語が子供にとって難解である場合，その語が検索の目的に沿っていて役に立つものであっても，子供が興味を持つことはなく，その語について調べないという状況も考えられる．さらに，検索エンジンを利用して検索結果を取得できても，ランキング形式の検索結果として提示されるWebページのタイトル，URL，Webページ中の一部の文章からなる情報では，どのページを選

[15] <Internet Explorer: http://www.microsoft.com/>

択するべきか迷ってしまい，ランキング1位のものから順に閲覧していくしかないかもしれない．

1.2.2項で述べたような，一般公開されている子供向けWebブラウザや子供向け検索エンジンも存在するが，基本的には子供にとって有害なコンテンツを除外するのみであり，それ以上の子供の特徴は考慮しておらず，一般向けのサービスと同じような問題がいまだ存在する．

1.2.1項で述べたような多くの既存研究にて，現在の子供のWeb閲覧・Web検索の際の問題点とそれらの問題点を解決するシステムの必要性が述べられている．特に，共通して指摘される問題点は，Web上の大多数の情報が子供にとって難解であり，子供はWebページを閲覧するにしろ，検索するにしろ，飽きやすく，すぐに目的から逸れてしまう傾向がある点である．そのため，大量の情報から，子供の興味をひくような子供にとって重要な情報に簡単に辿り着くことができない状況となっている．したがって，子供にとって快適なWeb閲覧・Web検索機構を実現するためには，子供にとって難解すぎず，興味を持てる情報を容易に取得可能とする必要があると考えられる．

さらに近年，1.2.2項，1.2.2項で述べたような，子供向けのWebシステムに関する研究に注目が集まっている．これらの研究では，子供が容易にWebを利用できるようにデザインされたシステムが提案されている．しかし，これらの研究では，大人が子供に見せたい情報のみに着目していたり，大人目線での使いやすさの評価にとどまり，子供による評価実験を今後の課題としている研究が多く，子供による評価実験を行っているものはほとんどない．そのため，子供にとって，本当に，興味を持つことができる形式で支援できているのかどうかは明らかになっていない．情報処理理論の分野では，内容を理解するためには，子供にとって楽しい経験を伴うことが重要であることが示唆されており [2]，子供のWeb閲覧・Web検索を支援する上で，子供の興味をひくかどうかという点は重要であると考えられる．

そこで本研究では，子供の興味をひくように，Web閲覧・Web検索の際に，子供が見たいものを中心に見せることに着目する．具体的には，自分が見たい情報にうまく辿り着けないような子供が，自分の興味をひく形式・内容の情報を容易に取得・閲覧できるように支援することを目的とする．子供の興味をひく情報を優先的に提供することで，子供は快適に楽しんでWebを利用することができるようになり，Webから情報を取得する意欲が高まり，子供がWebから様々な情報を学ぶきっかけ作りになることが期待される．

本研究では，1.2.1 項で述べた子供の特性に基づき，以下に示す特徴を満たす情報を子供の興味をひく，子供向けの情報として定義する．

- 構成

  文字ばかりで構成され，含まれている量が多すぎるような情報は，子供にとって面白みに欠け，注目すべきポイントが分からないと考えられる．そこで，子供にとっては，子供の興味をひき，分かりやすい構成が必要と考えられるため，画像やアニメーションが使用されており [58,59,73,74]，文章が適度な量であり [58,59,73,74]，カラフルな色使い [73,74] であるような情報を子供向けと定義する．

- 文章

  文章自体は，子供に親しみや興味を持たせるものであることが重要と考えられる．また，子供は学習のために Web を利用することが多いため[2]，子供にとって勉強に役立つ分かりやすい文章が必要と考えられる．そのため，長さは短めで，一文に含まれる情報が少なく [89]，漢字やアルファベットや難解な表現が少なく [58,59,89]，語りかけの表現などの子供向け特有の表現が含まれている [74,89] ような情報を子供向けとする．

以上のような定義に基づき，まず第 2 章では，子供が興味をひくページを容易に取得できるように，難解なページであっても子供が興味を持って閲覧できることを目指した Web ページ変換機構について検討する．ただし，元々子供にとって見た目にも内容的にも十分に分かりやすく，子供の興味を十分にひく，変換を必要としないような子供向けのページも存在する．しかし，子供は Web 検索の際に，検索結果から必要な情報を探すのを苦手とするという報告 [6] があるため，そのような子供向けのページに簡単にたどり着けるとは限らない．第 2 章の Web ページ変換機構は，このような問題を解決する 1 つのアプローチである一方，第 3 章では，Web ページの変換を行わない別のアプローチとして，子供の興味をひくページを優先的に提供する Web 検索機構について検討する．さらに，子供は Web 検索の際にクエリの入力を苦手とするという報告があるため，検索結果や Web ページの提示方法だけでなく，検索方法の支援も重要である．検索方法の支援を行う 1 つのアプローチとして，検索クエリに関するサブトピックを提示する方法が考えられる．クエリに関するサブトピックとは，入力されたクエリに対して “ユーザの意図を具体化する文字列” であり，例えば，“沖縄” というクエリについては，“歴史”，“食べ物” などがサブトピックの候

補として考えられる．文献 [65,66,69] では，子供のクエリ入力を支援するために，サブトピックを用いて検索クエリを拡張する方法について研究が行われているが，子供に提示すべき，子供が興味を持つサブトピックがどのようなものなのかは明らかではない．そこで第 4 章では，子供が興味をひく内容の情報を容易に検索できるように，子供の興味をひくサブトピックについての調査を行い，調査結果に基づき，検索クエリに関する子供向けサブトピックランキング手法について検討する．以下に各章の研究内容の概要を述べる．

1.3.2 泡メタファを用いた子供向け Web ブラウザ

Web ページの内容が理解できなければ，子供は興味を持てない．子供が興味を持って Web ページを閲覧するためには，見た目が興味をひくようなものであることが好ましく，その内容も子供が理解できる形態で表示する必要がある．

そこで本研究では，子供が Web ページを閲覧する際の支援方法の一つとして，泡メタファを用いたアニメーションによって，難解な Web ページを子供の興味を引く形式に変換する子供向け Web ブラウザについて検討する．これにより，子供が興味を持って，分かりやすく Web ページを閲覧することが可能となる．

1.3.3 Web 検索結果の子供向けリランク手法

上述のように，Web ページ自体を子供にとって興味をひきやすい形式に変換せずとも，元々十分に子供にとって見た目にも内容的にも分かりやすく，変換を必要としないような子供向けのページも存在する．そこで，Web 検索を行う際に，そのような子供向けの Web ページを容易に取得可能とするための検索機構について検討する．

Web 検索を行う場合，自分が必要としている情報をランキング形式で表示される検索結果の中から選択する必要がある．しかし，一般的な検索エンジンはもちろん，子供向けの検索エンジンでも，基本的に検索結果のランキングは子供向けでない．例えば，一般的な検索結果において，Wikipedia[16] の記事が上位であることが多いが，このようなページは文章が多く，難解な表現が多用されているため，子供の興味をひく分かりやすいページとはいえない．また，子供向け検索エンジンにおいても，あらかじめ登録されたお勧めサイト以外のランキングは一般的なものと同じであり，お勧めサイトであっても必ずしも子供

[16] <Wikipedia: http://ja.wikipedia.org/wiki>

向けであるとはいえない．そのため，現状の検索エンジンでは，子供の興味をひくような情報がランキングの上位であるとは限らない．子供は，検索結果から自分が求めている結果を選択するのが苦手であり，検索結果の上位5件程度しか閲覧しないという特徴があるため [16]，検索結果のランキングでは，子供向けのページを上位にランクする必要がある．子供向けの検索結果のランキングに着目した研究 [34,88] もあるが，子供にとって興味をひくWebページであるのかどうかについては直接考慮していない．

そこで本研究では，子供の興味をひくページを上位にランクする検索結果の子供向けリランク手法を提案する．この手法により，子供向けのページほど検索結果の上位にランクされるため，子供はランキング下位まで探す手間なしに，子供向けのWebページを容易に得ることができる．

1.3.4 Web検索クエリに関する子供向けサブトピックランキング手法

子供はWeb検索の際にクエリの入力を苦手とするという報告 [16] があるため，検索結果やWebページだけでなく，検索方法の支援も重要である．子供のための検索方法の支援として，文献 [56,65,66,69,78] では，オントロジによるクエリ拡張や画像を用いたクエリ入力といった支援が行われている．しかし，これらの手法で提示されるクエリに関するサブトピックが，本当に子供にとって興味をひき，重要なものなのかは明らかではない．

子供の興味をひく重要なサブトピックは，子供の検索目的を考慮すると，“趣味” や “学習” に役立つものであると考えられる．例えば，クエリ “沖縄” については，“ツアー”，“レンタカー” というような旅行に関するサブトピックが一般的に重要と考えられるが，子供自ら沖縄旅行をする際に，ツアーやレンタカーは子供にとって興味をひく内容ではなく，それらに関する情報を調べる状況は考えにくい．一方，“食べ物”，“歴史” というようなサブトピックは，子供が授業で沖縄の文化や習慣について学んでいる際に興味をひく内容であり，重要であると考えられる．子供の興味をひくサブトピックを提示することで，そのサブトピックについて調べてみたいと子供の欲求を刺激し，Webを活用するきっかけになると考えられる．しかし，1.2.2項で述べた子供のWeb検索行動を調査する既存研究では，子供の興味をひく，子供にとって重要なサブトピックについては調査が行われていない．

そこで本研究では，子供の興味をひくサブトピックがどのようなものなのか，どのように抽出すべきなのかについての調査を行う．さらに，その調査結果に基づき，子供向けサブトピックを優先的に提供するためのサブトピックランキング手法について検討する．こ

れにより，子供は検索クエリに関するサブトピックを入力するという手間なしに，子供向けサブトピックに関する情報を容易に得ることができる．

1.4 本論文の構成

本論文は，5 章から構成され，その内容は次の通りである．まず，第 2 章では，泡メタファを用いた子供向け Web ブラウザを提案する．第 3 章では，Web 検索結果の子供向けリランク手法を提案する．第 4 章では，Web 検索クエリに関する子供向けサブトピックランキング手法を提案する．最後に第 5 章で，本論文の結論と，今後の検討課題について述べる．なお，第 2 章は，文献 [46,48,50,52] で公表した内容に基づき論述する．第 3 章は，文献 [45,47,49,51,54] で公表した内容に基づき論述する．第 4 章は，文献 [53] で公表した内容に基づき論述する．

第2章

泡メタファを用いた子供向けWebブラウザ

2.1 まえがき

第1章で述べたように，子供にはWeb閲覧に関連する以下のような特徴があることが従来の調査などで知られている．

(1) 全体に散らばった情報を楽しんで探す [82].

(2) 理解を深くするには見せたい情報のみを与える [57].

(3) 楽しい経験は学習に繋がる [2].

(4) キャラクタなどのイラスト中心のページに興味を持つ [58].

(5) 難しい漢字，長い文章のページを苦手とする [58].

(6) カテゴリ概念は苦手だが，具体例があれば理解できる [87,90].

現状ではこのような大人とは異なる子供の能力や嗜好を考慮したWebブラウザは存在せず，子供がWebページを閲覧するには，Webページの内容の理解が困難である，表示形態が面白みにかけるといった問題点があると考えられる．

例えば，多くのWebページでは子供にとって難解な表現や漢字が使用されている．難解な表現，漢字が使われることで子供は文章を読むことができず，画像やWebページの雰囲気からしか内容を知ることができない．また，小学生の国語の教科書に載っている文章は短文であり，大きな文字で書かれているが，一般のWebページでは，文章が長く，文字が小

さいことが内容の理解を妨げている一因といえる．さらに，一般向けページの1つの Web ページに表示されるコンテンツ数は子供向けページより多く，子供にとっては全ての内容を一度に把握しきれないと考えられる．また，ニュースサイトのようなページは大部分が文章で構成され，画像が少なく，色使いが単調な構成である．このような Web ページでは，子供が閲覧途中であきてしまい，最後まで Web ページを閲覧するだけの興味を持続できないものと考えられる．

当然，Web ページの内容が理解できなければ，子供は興味を持てない．子供が興味を持って Web ページを閲覧するためには，見た目が興味をひくようなものであることが好ましく，その内容も子供が理解できる形態で表示する必要がある．情報処理理論の分野では，内容を理解するためには，子供にとって楽しい経験を伴うことが重要であると示唆されている [2]．しかし，上記のような問題点があるため，インターネットは子供に学ぶ機会を与えてくれる場であるにも関わらず [28]，十分にインターネットを活用することができていない．

そこで本章では，子供の Web 閲覧能力や嗜好を考慮して見ていて楽しくわかりやすい Web ページ提示を実現するための1つのアプローチとして，Bubble Browser を提案する．Bubble Browser では，Web ページのコンテンツを泡メタファを用いて表示することで，コンテンツの情報を泡のスピードや色，サイズなどに反映し，わかりやすい視覚的な情報として伝える．さらに，コンテンツの詳細内容は絵本形式に変換して提示することで，子供が興味を持ち，理解しやすいコンテンツ表示を実現する．これにより，子供が様々な一般向け Web ページを楽しみながら閲覧でき，インターネット上の有益な情報を得るという学習に結びつく．

以下では，2.2 節で提案インタフェースである Bubble Browser の設計と実装について述べる．その後，2.3 節で実施した評価実験について述べ，2.4 節で本章のまとめを行う．

2.2 Bubble Browser の設計と実装

2.2.1 設計方針

2.1 節で述べた子供の特徴に基づき，子供向け Web ブラウザの設計方針を以下のように設定した．

(1) 複数の情報を散らばらせた表示

子供は，全体に散らばった情報を見回して欲しい情報を探すことを楽しむという検索性を持つため，Web ページの情報についてもその検索性を満たす表示方法にすることが有効である．具体的には，Web ページ中の複数のコンテンツ情報を画面上に散らばらせて表示することで，画面全体から欲しい情報を探すことができる機能が効果的であると考えられる．ただし，情報量は子供にとって負担にならない程度でなければならないので，Web ページ中の全ての情報を表示するのではなく，ある程度要約したような簡潔な情報とした方が子供に適している．

(2) 表示する情報の絞り込み

子供が内容を深く理解するためには，与える情報を少なくする必要があるので，表示する情報を絞り込む機能が有効である．そのため，始めは多量の情報から興味のある情報を能動的に探し，その後，興味のある情報のみの表示とする表示方法が有効であると考えられる．

(3) 見ていて楽しい表示

子供には学習に楽しい経験が必要であり，キャラクタなどのイラスト中心の Web ページに興味を持ちやすいことが分かっている [58]．そこで，画像，イラストやアニメーションを用いた，色使いを多くしたカラフルな表示が有効である．

(4) 分かりやすく親しみやすい形式のテキスト情報

子供は，難しい漢字，長い文章が用いられた Web ページが苦手ですぐに飽きてしまう傾向があるため [58]，テキスト情報は分かりやすく親しみのある形式へ変換する機能が有効である．具体的には，文章を短く区切り，難しい表現は容易な表現に変換することが有効である．

(5) カテゴリごとの具体例の表示

子供は，カテゴリ概念を苦手とするため [83]，Web ページにおける一般的なカテゴリ分けは理解が困難である．そこで，具体例を提示することでカテゴリ概念の理解の助けになるため [87,90]，そのような機能が有効であると考えられる．

以上の設計方針を満たすように，提案ブラウザでは泡メタファを用いた表示を行う．ここで，泡メタファを採用したのは，泡を画面上に散らばらせて表示し，アニメーションにす

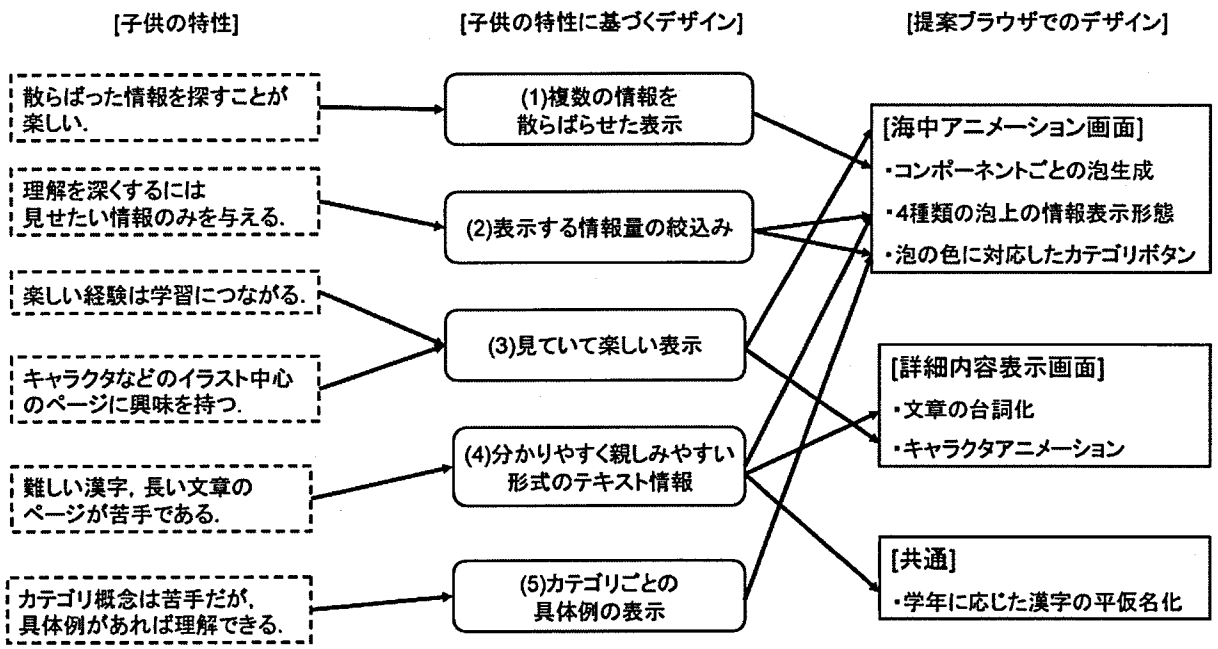


図 2.1: 子供の特性とデザインの対応関係

ることが可能なためである．また，泡上に情報を表示することが容易であり，さらに，泡の色，サイズ，スピードなどにより視覚的に情報の属性を表現できる．

具体的には，図 2.1 に示すような対応関係で提案ブラウザを設計した．方針 (1) を実現するため，Web ページをコンポーネントという意味的なブロックに分割し，コンポーネントごとに生成した泡を画面上に散らばらせて表示する．コンポーネントについては 2.2.3 項で説明する．また，方針 (2) を実現するため，簡潔な情報のみを泡上に表示し，表示する泡をユーザが絞り込めるようにする．方針 (3) を実現するため，Web ページは海中アニメーションへ変換し，泡に含まれたコンポーネントの詳細内容は，キャラクタを用いた絵本形式へ変換する．そして，方針 (4) を実現するため，アニメーション画面でのテキスト情報は簡潔にし，コンポーネント内の詳細な文章は台詞へ変換する．このとき，テキスト情報中の漢字は平仮名に変換する．最後に，方針 (5) を実現するため，カテゴリごとの代表語を表示できるようにする．

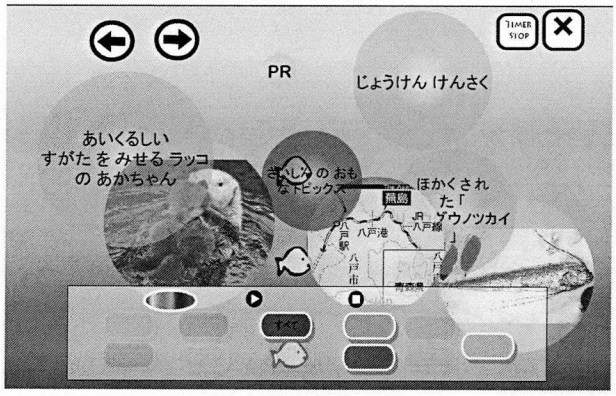


図 2.2: 海中アニメーション画面

図 2.3: 詳細内容表示画面

2.2.2 基本動作

提案ブラウザは海中アニメーション画面で Web ページの概要情報を表示し，その後，興味を持った泡に該当するコンポーネントの詳細内容を表示する．このとき全てのテキストに含まれる漢字はルビ振り API[1] を用いて学年に応じた平仮名に変換する．具体的には，以下のような手順を繰り返し Web ページを提示する．

1. 海中アニメーションにより Web ページを提示する．(図 2.2)
2. ユーザはページ全体の情報を閲覧し，興味を持った泡をクリックする．
3. 泡に含まれていたコンポーネントの詳細内容を絵本形式で提示する．(図 2.3)
4. ユーザは詳細内容を閲覧する．
5. ユーザが興味を持ったリンクをクリックすると，リンク先ページの取得，解析を行い，リンク先ページに対する海中アニメーションを表示する．

またユーザは，画面左上の矢印ボタンで通常の Web ブラウザと同様のバック，フォワード操作を行える．

[1] <Yahoo!デベロッパーネットワーク ルビ振り API: http://developer.yahoo.co.jp/webapi/jlp/furigana/v1/furigana.html>

図 2.4: コンポーネントの例

2.2.3 機能の設計

海中アニメーション画面

海中アニメーション画面では，図 2.2 に示すように，コンポーネントごとに生成した泡が様々なサイズ，上昇スピードで海底（画面下部）から海面（画面上部）に向かって上昇する．コンポーネントとは，Web ページ中の関連のある情報が集まったブロックのことである．例えば，Yahoo! JAPAN などのポータルサイトに見られる検索フォーム，サイト内の各ページへのリンクの集合であるメニューリストや，ニュース，ディレクトリなどがコンポーネントである．図 2.4 は Yahoo!JAPAN の Web ページに含まれるコンポーネントを示したものである．なお，本研究で用いるコンポーネント分割手法は，筆者の所属する研究チームが提案した方式 [62] を用いる．コンポーネントごとに泡にすることで，ユーザは関連する情報ごとにまとめて効率的に閲覧することができ，興味をもった情報を容易に選択，閲覧できる．

泡のサイズや色は，コンポーネントの情報量やカテゴリなどの特徴により決定し，泡上にはコンポーネントに含まれている情報を表示する．泡が出てくる順序については，Web ページにおいて左上に位置するコンポーネントから右下に位置するコンポーネントへと順に出現し，海面に達した泡は循環して海底から再び現れるものとする．泡はユーザが自由にドラッグして動かすことができ，カーソルが泡の中に入ったときは泡を拡大して表示し，ユーザが注目した泡上の情報を見やすくする．具体的には，以下のような機能を実現する．

- 情報量に対応した泡のサイズと上昇スピード

  一般的に，コンポーネント内の情報量が大きいものほどユーザにとって有益なコンテンツであると考えられる．ここで，情報量とはコンポーネントに含まれる文字数と画像数と定義する．Bubble Browser では，コンポーネントの情報量と泡の大きさを対応させ，情報量の大きいものほど泡のサイズを大きくし，ユーザの関心をひくようにする．つまり，情報量の小さいヘッダーやフッターといったコンポーネントは小さな泡となり，テキストやリンクを多数含むメインのコンポーネントは大きな泡となる．また，情報量の小さいものほど，泡の上昇スピードを大きくする．これにより，情報量の大きな泡ほど画面上に長時間滞留し，ユーザの目に触れやすくなる．

- 4 種類の泡上の情報提示形態

  コンポーネントの内容を簡潔に表すため，泡上に表示する情報は以下に示すように 4 種類とし，ユーザがボタン 1 つで簡単に切り替えられるようにする．

  - タイトル＋関連画像：概要を把握するためのデフォルト表示．(図 2.5(a))
  - 関連画像：視覚性を重視するための画像情報を主とした表示．(図 2.5(b))
  - タイトル＋代表語：テキスト情報を付加した表示．(図 2.5(c))
  - コンポーネント画像:コンポーネントのスクリーンショット画像の表示．(図 2.5(d))

  タイトルとは，コンポーネントの内容を端的に表す “トピックス”，“おすすめセレクション” などの 1 フレーズもしくは 1 文とする．コンポーネントのテキストのうち，強調表現が使用されている部分が内容を表すカテゴリ，記事のタイトルなどである場合が多い．そこで，タイトルについては強調 HTML タグに含まれるテキストを抽出する．強調 HTML タグとは，$\langle strong \rangle$，$\langle h1 \rangle$ タグのようなテキストの強調表示に用いられるタグである．このとき，抽出したテキストが複数文で構成される場合は，句点でテキストを切り，最初の 1 文のみを用いる．強調 HTML タグがコンポーネントのテキストに存在しない場合は，最初の 1 文が記事全体の内容の要約になっていることが多いため，テキスト全体の最初の 1 文を用いる．

  代表語とは，コンポーネントの内容を表す代表的な語であり，コンポーネントのテキスト中の名詞に重み付けを行うことで抽出する．具体的には，まずコンポーネントのテキストについて形態素解析を行い名詞を抽出する．このとき，連続する名詞は複合

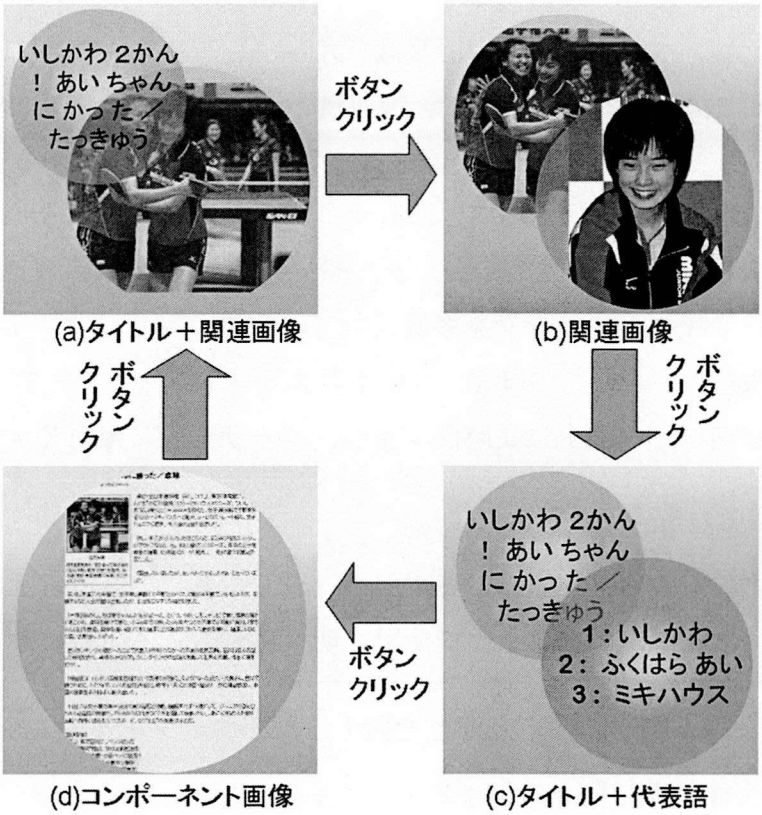


図 2.5: 情報表示形態の例

語として1つの名詞とする．そして抽出した名詞に重み付けを行い，重みの大きい3語を代表語として選択する．一般的に，固有名詞や未知語はコンポーネントの内容を表す特徴的な語の1つであるため，まず語の属性による重み付けを行う．また，強調HTMLタグに含まれる語の重要度も高いと考えられるため，強調HTMLタグによる重みも付加する．さらに，コンポーネント内に複数回現れる語は特徴的な語の1つであると考えられるため，出現回数も考慮する．

関連画像とは，コンポーネントの内容に関連する人物の写真などの画像とする．これには，コンポーネント内の画像と画像検索API[2] で検索した結果の画像を抽出して用いる．具体的には，まずコンポーネント内に画像が存在する場合は，その画像を抽出する．コンポーネント内には記事の内容を表すような画像もあれば，アイコン画像，メニュー画像やデザイン画像などコンポーネントの内容とは直接関係のない不要な画

[2]<Yahoo!デベロッパーネットワーク 画像検索 API: http://developer.yahoo.co.jp/webapi/search/imagesearch/v2/imagesearch.html>

像も多く存在する．このような画像を除外するため，文献 [41,61] を参考に，JPEG 画像以外の縦，横の両方が 30[pix] 未満の画像，1 つの Web ページ中に繰り返し出現する画像は除外する．

コンポーネント内の画像が 2 個に満たない場合のみ，画像検索を用いて画像を抽出する．画像検索を行う検索語は，代表語の抽出方法と同様にコンポーネントのテキストから抽出した名詞に重み付けをすることで決定する．そして，重みの大きい語を用いて画像検索し，検索結果 1 位の画像を関連画像とする．

- コンポーネントのカテゴリによる泡の色分けとカテゴリボタン

小学生のインターネットの利用目的は，“趣味や遊びのことを調べる”，“勉強のことを調べる” という 2 つに大きくわけられる[3]．そこで，Web ページのコンポーネントを “趣味”，“勉強” という 2 つのメインカテゴリに分けて提示することで，子供にとって必要な情報が見つけやすくなると考えられる．代表的な子供向けポータルサイトである Yahoo!きっず，キッズ goo のカテゴリを参考に，趣味，勉強の 2 つのカテゴリにサブカテゴリを設定した．趣味については 5 つのサブカテゴリにわけ，勉強については 4 つのサブカテゴリにわけた．そしてサブカテゴリごとに泡の色を変えることで，視覚的にカテゴリを区別できるようにする．また，一目でメインカテゴリが区別できるように趣味には暖色系，勉強には寒色系の色を用いる．図 2.6 にそれぞれのサブカテゴリとサブカテゴリに応じた泡の色を示す．

コンポーネントのカテゴリ分けは，カテゴリごとに関連語を記述した辞書を用いる．まず，Yahoo!きっずやキッズ goo のカテゴリを参考に，各カテゴリごとにベースとなる関連語を定義する．そして，インターネット上のフリー百科事典である Wikipedia から構築された Wikipedia シソーラス [70] から，Wikipedia API[4] を用いて，ベースとなる関連語から連想される語をさらに 30 語ずつ抽出し，辞書とする．Wikipedia には既存の辞書には含まれていない新語や専門語，商品やサービス名も含まれているため，多様な情報が含まれる Web ページのコンポーネントの分類に有効であると考えられる．以上の方法を用いて各カテゴリの辞書を構築し，コンポーネントのテキスト中にどのカテゴリの辞書の語が最も出現するかでカテゴリを判定する．辞書とマッチ

[3]<NTT レゾナント : http://research.goo.ne.jp/>

[4]<Wikipedia API: http://wikipedia-lab.org/>

(a)ゲームと遊び: 濃いピンク (b)アニメと漫画:赤 (c)スポーツと趣味: オレンジ (d)テレビ:薄いピンク (e)ニュースとメディア: 黄

(f)理科:薄い青 (g)社会:紫 (h)言語:濃い青 (i)その他の科目: 水色 (j)その他:緑

図 2.6: カテゴリと泡の色

した語の数が 2 語以下の場合は，どのカテゴリにも属さない“その他”のカテゴリとする．

さらに Bubble Browser では，カテゴリボタン機能を提供する．カテゴリボタンは図 2.7 のように，以下の 2 つの機能を提供する．

- 泡の絞り込み：ユーザがカテゴリボタンを押すと，選択カテゴリに属するコンポーネントの泡のみを表示する．選択された泡は画面中央に 1.2 倍に拡大して表示する．
- 代表語の表示：ユーザがカテゴリボタン上にカーソルを乗せると，選択カテゴリに属するコンポーネントの代表語を 5 語表示する．

詳細内容表示画面

詳細内容表示画面では，子供がテキストを中心として構成された Web ページでも飽きずに閲覧できるように，絵本形式 [85] に変換して表示する．

Web 上で閲覧できる従来の絵本は，上部に大きくイラスト，もしくは，アニメーションを表示し，下部に文章を表示する形式となっている．文章は登場人物同士の会話やナレーターによる説明により構成される．そこで，Bubble Browser では，2 匹の魚のキャラクタアニメーションとキャラクタの会話形式による絵本に変換する．図 2.8 のように，ユーザ

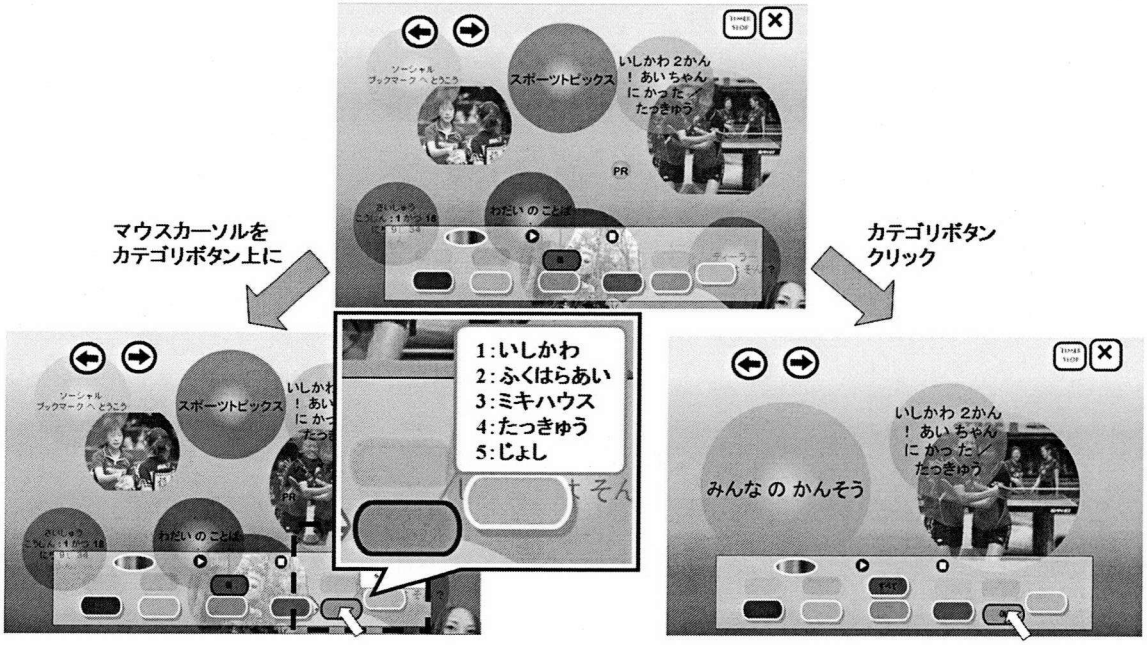


図 2.7: カテゴリボタンの動作例

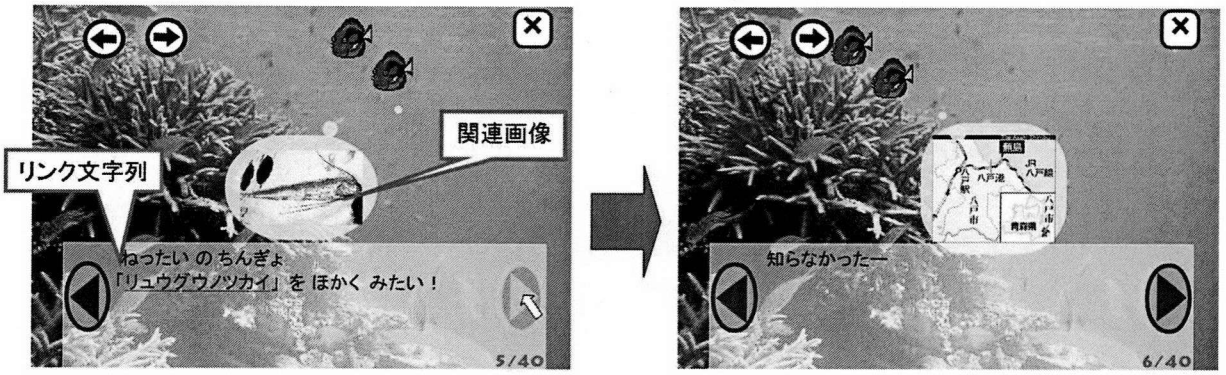


図 2.8: 絵本による詳細内容表示画面

は画面下の青い右矢印ボタンを押すことで次の文章，左矢印ボタンで前の文章を閲覧でき，会話の文章内のリンク文字列を選択することでリンク先ページに移動できる．また，キャラクタのふきだし中にコンポーネントの関連画像を一定時間ごとに順番に表示することで，コンポーネント内容の理解を補助する．

キャラクタについてはあらかじめ用意しておいた10種類の画像からランダムに2種類使

用し，会話は文献 [39] を参考に，コンポーネントの各文章の末尾に語尾，応答を付加して作成する．具体的には，まずコンポーネントの文章を HTML の ⟨*br*⟩ タグで分割し，さらに句点，改行コードで分割する．このようにして得た各文章に対して，語尾テンプレート，応答テンプレートからランダムに語尾，応答を抽出し，各文章の末尾に付加する．語尾テンプレートには「らしいよ！, みたい！, ってさ！, らしい, だってさ♪, なんだって, みたいだよ, らしいねー, だねー, だって」，応答テンプレートには「へぇー!!, そうなんだ♪, へー！そうなんだね！, ふーん, そっかぁ, そっか！すごいね♪, すごいや!!, 知らなかったー, そうなの??, もっと聞かせて！」のそれぞれ 10 種類の語尾，応答をあらかじめ用意した．

システムの実装

Windows XP 搭載の PC 上で，Visual Studio.NET Visual C#言語で Web ページの解析部分を，Adobe Flash CS3 Professional でアニメーションの表示部分を実装した．解析部分では，Web ページからコンポーネントを抽出し，各コンポーネントの特徴を決定する．なお，コンポーネントのテキストに含まれる語の品詞の判別には MeCab[5] を用いた．そして，コンポーネント情報，コンポーネント特徴情報に基づいたアニメーションを Flash を用いて作成し，ユーザに提示する．

2.3 評価実験

本節では提案ブラウザの有効性を確認するために行った評価実験について述べる．

2.3.1 実験環境

被験者は 4 才から 10 才の 13 名の男女とし，その内訳は，4 才の女子 1 名，5 才の男子 1 名，6 才の男子 2 名，7 才の女子 1 名，8 才の男子 1 名，9 才の男子 4 名，10 才の男子 2 名，女子 1 名である．本実験では，2.2 節で述べた提案ブラウザである Bubble Browser，さらに，比較対象として一般向け Web ページ中の漢字を学年に応じた平仮名に変換して提示する “ひらがなブラウザ”，子供向け Web ページを元の形態のまま提示する “一般ブラウザ” の 3 種類のブラウザを用いて，4 つの Web ページを閲覧してもらった．

[5] <MeCab: http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

表 2.1: 小学校入学前グループ用の質問一覧表

	タイプ1	タイプ2
一般ページ1	何のスポーツが出てきましたか？	勝ったのは誰でしょうか？
一般ページ2	何の動物が出てきましたか？	その動物は何をしていたでしょうか？
子供ページ1	何の動物が出てきましたか？	その動物はどれだけいましたか？
子供ページ2	誰が出てきましたか？	それはどこの大統領ですか？

Web ページは Yahoo! ニュース[6] の一般向けページ 2 種類，Yahoo!きっずニュース[7] の子供向けページ 2 種類とした．Bubble Browser で一般向けページ 1 つ，ひらがなブラウザで一般向けページ 1 つ，Bubble Browser で子供向けページ 1 つ，一般ブラウザで子供向けページ 1 つの 4 ページを閲覧してもらった．これは，一般向けページと子供向けページそれぞれにおいては Bubble Browser を用いる効果が異なると考えたためである．

そして，学習効果を検証するため Web ページを閲覧中に，ページの内容について 2〜4 問の質問に答えてもらった．質問は，海中アニメーション画面の効果，詳細内容表示画面の効果のそれぞれを調べるため，2 種類を用意した．海中アニメーション画面では，簡潔な情報を表示することで Web ページ中の概要を楽しみながら把握してもらうことを目的としているため，概要を把握しているかを確認するための質問（タイプ 1）1 個に答えてもらった．一方，詳細内容表示画面では，詳細内容の楽しみながらの閲覧を学習に繋げることを目的としているため，詳細内容を理解したかどうかを確認するための質問（タイプ 2）を 1〜3 個実施した．ただし，年齢によって理解能力が異なると考えられるため，小学校入学前用，小学校低学年用，小学校中学年用の 3 つの種類の質問を用意した．タイプ 1 の質問は，小学校入学前の被験者のみ，絵から答えを選択する 3 択問題とし，小学校低学年以上の被験者は口答で答える問題とした．タイプ 2 の質問は，小学校入学前の被験者は絵から答えを選択する 3 択問題 1 問，小学校低学年の被験者は絵から答えを選択する 3 択問題 1 問，日付などの 3 択問題 1 問の計 2 問，小学校中学年の被験者は絵から答えを選択する 3 択問題 1 問，日付や場所などの 3 択問題 2 問の計 3 問とした．年齢に応じた各ページに対する質問を表 2.1，2.2，2.3 に示す．

[6] <Yahoo! ニュース: http://headlines.yahoo.co.jp/hl>

[7] <Yahoo!きっずニュース: http://news.kids.yahoo.co.jp/>

表 2.2: 小学校低学年グループ用の質問一覧表

	タイプ1	タイプ2
一般 ページ1	何のスポーツが出てきましたか？	勝ったのは誰でしょうか？ 写真の人のユニフォームは何色でしょうか？
一般 ページ2	何の動物が出てきましたか？	その動物は何をしていたでしょうか？ いつまで見られるでしょうか？
子供 ページ1	何の動物が出てきましたか？	その動物はどれだけいましたか？ その動物は去年9月に何をされましたか？
子供 ページ2	誰が出てきましたか？	それはどこの大統領ですか？ 就任式はいつですか？

表 2.3: 小学校中学年グループ用の質問一覧表

	タイプ1	タイプ2
一般 ページ1	何のスポーツが出てきましたか？	勝ったのは誰でしょうか？ 写真の人のユニフォームは何色でしょうか？ 記事の人は何を目指していますか？
一般 ページ2	何の動物が出てきましたか？	その動物は何をしていたでしょうか？ いつまで見られるでしょうか？ ミュージアムは何県で見られますか？
子供 ページ1	何の動物が出てきましたか？	その動物はどれだけいましたか？ その動物は去年9月に何をされましたか？ それは最初にどこで見られましたか？
子供 ページ2	誰が出てきましたか？	それはどこの大統領ですか？ 就任式はいつですか？ コンサートがあったのはどこですか？

2.3.2 実験手順

以下のような流れでWebページを閲覧してもらった.

1. Bubble Browserでは海中アニメーション画面，比較ブラウザではWebページ内のタイトル，画像などの概要を軽く閲覧してもらう.

2. タイプ1の質問1問に答えてもらう.

3. Bubble Browsr では詳細内容表示画面，比較ブラウザではコンポーネント内の画像，詳細な文章をじっくり閲覧してもらう.

4. タイプ2の質問1〜3問に答えてもらう.

5. 1〜4の手順を4つのページで繰り返す.

このとき，実験時間の制約上の問題から，小学校入学前の被験者13人のうち3人にはBubble Browserで2ページのみ閲覧してもらい，ひらがなブラウザ，一般ブラウザでの閲覧は行わなかった．実験では，被験者が用いるブラウザと閲覧するページの組が固定されないように，ブラウザと閲覧するページの組合せはランダムに決定した.

全てのページの閲覧後，使用したそれぞれのブラウザに対し，楽しいかどうかについて3段階（楽しい，普通，楽しくない），今後使ってみたいかどうかについて2段階（使いたくない，使いたい），操作が難しいかどうかについて3段階（難しい，普通，簡単）で評価してもらった．さらに，小学校低学年以上の被験者には，泡の数，サイズ，色の種類，速さ，表示する画像や言葉のわかりやすさ，詳細内容表示画面の楽しさについても3段階で評価してもらい，泡に表示させてほしい情報について自由形式で記入してもらった.

2.3.3 実験結果

質問に対する正答率

質問に対する正答率の結果を表2.4に示す．表2.4からわかるように，タイプ1の質問の正答率は一般向けページではBubble Browserがひらがなブラウザを大きく上回り，子供向けページではどちらのブラウザも同じ結果となった．これは，一般向けページは子供にとっては情報過多で一見しただけでは内容を把握しにくいが，Bubble Browserの海中アニメーション画面ではタイトル，関連画像のみの簡潔な表示を実現しているために概要を把握しやすかったためであると考えられる．一方，子供向けページは元々子供にとってわかりやすい情報量であるため，Bubble Browserを使わなくても概要を把握するのが容易であったことが分かる.

タイプ2の質問の正答率も一般向けページではBubble Browserがひらがなブラウザを大きく上回った．これは，一般向けページでは文章表現が難解であるのに対して，Bubble

表 2.4: 各質問に対する正答率

		タイプ1 (海中アニメーション画面)	タイプ2 (詳細内容表示画面)
一般向けページ	Bubble Browser(全13人)	100%	90%
	ひらがなブラウザ(全10人)	80%	69%
子供向けページ	Bubble Browser(全13人)	100%	76%
	一般ブラウザ(全10人)	100%	81%

Browserでは文章以外の情報が内容理解に大きく影響したものと考えられる．つまり，Bubble Browserの詳細内容表示画面では，ひらがなブラウザで表示されない関連画像が表示されるため，理解の補助に繋がったと考えられる．一方，子供向けページでは，タイプ2の質問の正答率はBubble Browserが若干下回っていた．これは，子供向けページが容易な文章表現であり，一般向けページとは異なり，文章からのみでも内容の理解が容易であったためと考えられる．つまり，画像がなくても文章自体が子供向けであれば，小学校中学年程度ならばそれを熟読することで内容を理解できたものと考えられる．Bubble Browserの正答率が低くなってしまったのは，元々子供向けに作られた丁寧な口調の文章を会話形式に変換したことで，文章の文末表現が長くなり，むしろ分かりにくくなってしまったものと考えられる．例えば,「しないのすぎのきにとまったトキ3わがもくげきされました.」のような文章を変換した場合,「しないのすぎのきにとまったトキ3わがもくげきされましたらしいよ♪」となり，不自然な文章となってしまったケースがあった．したがって，文章を会話形式に変換する際には，自然な会話になるように文末を考慮するなどの工夫が必要である．

ここで，タイプ1の質問の正答率に対してタイプ2の質問の正答率が全体的に低くなっているのは，タイプ2の質問がタイプ1の質問に比べて難易度が高かったためであると考えられる．詳細内容表示画面では被験者がコンポーネントの文章全てを閲覧した後に質問するため，細かい数字などの詳細内容を記憶しておかなければ正答することができない．また，誤答を招きやすい3択問題も実施しており，例えば，子供向けページ1の「それは最初にどこで見られましたか？」という質問に対しては,「水田，民家，杉の木」という全て

文章内に現れる語による3択にした．この質問の正解は「杉の木」であるが，文章内の最後に出てきた「水田」と答えた被験者が多かった．このような3択の質問は特に正答率が低かった．したがって，細かい内容までよりわかりやすく提示するためには，表示文章に応じたイラスト，アニメーション提示などの視覚的な補助が必要であると考えられる．

アンケート結果

各ブラウザを使ってみて楽しかったかどうかのアンケート結果を図2.9に示す．ひらがなブラウザ，一般ブラウザでは楽しいと答えた被験者は33%であったのに対し，Bubble Browserでは83%の被験者が楽しいと答えた．また，各ブラウザを今後使いたいかどうかのアンケート結果を表2.5に示す．ひらがなブラウザ，普通ブラウザでは33%の被験者が使いたくないと答えたが，Bubble Browserでは100%の被験者が使ってみたいと答えた．これらの結果から，Bubble Browserは，カラフルなアニメーションを使うことで，一般的なブラウザ，ひらがなに変換したのみのブラウザよりも，最初の印象として，子供にとって楽しく，再び使いたくなるような提示を行うことができたと考えられる．実際に，タスク終了後も2名の被験者がBubble Browserを使ってWeb閲覧を続けていた．ただし，文字ばかりの元々のWebページを提示するよりも，アニメーションを利用して提示する方が子供にとって楽しい結果となることは明らかであり，長期間使用した際にもその効果が持続するかどうかについても，検証する必要がある．

さらに，各ブラウザが使いやすかったかどうかのアンケート結果を図2.10に示す．各ブラウザで大きな差は見られなかったが，Bubble Browserは操作が簡単と答えた被験者は，ひらがなブラウザより11%少ない結果となった．これは，Bubble Browserはひらがなブラウザよりボタン操作が多いためであると考えられる．しかし，他のアンケート結果が示すように，Bubble Browserは楽しい，使いたいという意見の被験者が多く，操作性が被験者のストレスには繋がらなかったため，操作性の問題はそれほど大きくないと考えられる．

被験者の意見

Bubble Browserのデザインについて，質問内容が理解できると考えられる8才以上の被験者8人に意見を聞いたところ，“泡の数が少ない”と答えた被験者が1人，“泡のサイズが大きい”と答えた被験者が3人，“泡の色の種類が多い”と答えた被験者が2人，“少ない”と答えた被験者が1人，“上昇スピードが遅い”と答えた被験者が3人，“速い”と答えた被験

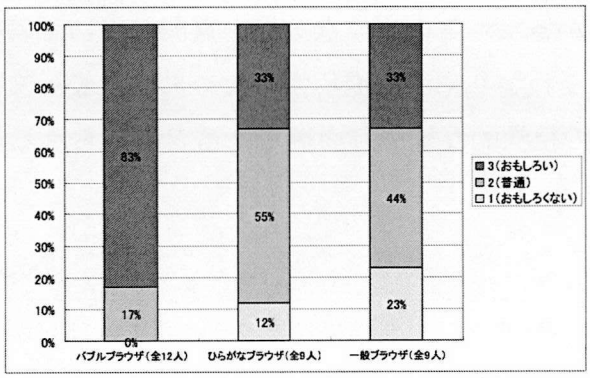


図 2.9: 各ブラウザの楽しさ

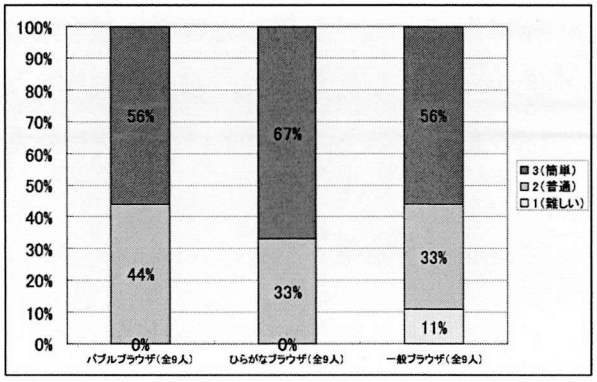


図 2.10: 各ブラウザの操作性

表 2.5: 各ブラウザの反復性

	使いたい	使いたくない
バブルブラウザ（全11人）	100%	0%
ひらがなブラウザ（全9人）	67%	33%
一般ブラウザ（全9人）	67%	33%

者が1人いた．また，“泡に表示している画像やテキストがわかりにくい”と答えた被験者が2人いた．その理由として，テキストが長くてわかりにくいという意見を得た．さらに，“絵本による表示が楽しい”と答えた被験者は8人全員だった．その他に“絵本ならニュースなども読んでみようという気になる”，“画像を多く表示させて欲しい”，というような意見を得た．これらの意見から，泡にはもっとわかりやすいイラスト，画像を表示させ，テキストについてはタイトルをもっと短く，簡潔にする必要があると考えられる．見やすい泡の数，サイズ，スピードは個人差があるようなので年齢に応じて設定を変化させる，もしくは，ユーザ自身で調節できるようにする必要がある．また，詳細内容をアニメーションと会話形式の絵本に変換することで，ニュースサイトのようなわかりにくいページでも親しみやすさが増し，子供にとって楽しい表示が実現できたと考える．特に，低年齢の被験者ほど会話形式の文章を楽しんで読んでいる様子が見られたことより，文章を会話形式に変換する方法は文章を読むことに慣れていない子供には有効であると考えられる．

2.3.4 考察

海中アニメーション画面

Bubble Browserにおいて，現在は泡に表示する情報を4種類としているが，実験の結果，コンポーネントの内容に関連する写真，イラストやアニメーションを中心とした表示に変更し，テキスト情報は補佐的な役割とする必要があることがわかった．しかし，現在の方法では，画像検索に用いるクエリの精度が十分でないため，コンポーネントの内容に無関係な画像も抽出されることがある．そこで，画像検索に用いるクエリの精度を上げ，検索結果から正しい結果のみを抽出する必要がある．もしくは，リンク先ページの画像を用いる方法も考えられる．また，現在のカテゴリの定義ではあらかじめカテゴリの種類を限定しているため，ページによっては同じ色の泡のみになってしまう．そこで，カテゴリの粒度を細かく定めておき，ページによって適切な粒度のカテゴリを用いることで，同じ色の泡のみになってしまわないようにする必要があると考えられる．以上のような点を改善し，海中アニメーション画面をより見やすく楽しいものとすることが重要である．

詳細内容表示画面

現在の実装では，あらかじめ用意しておいたキャラクタ，語尾，応答を付加する方法で絵本化を行っているが，文章内の難解な表現を辞書などを用いて子供向けの表現に変換し，難解な単語については意味を調べる機能を付加し，文章を容易にする必要がある．さらに関連画像をふきだしに表示するだけでなく，楽しい話題，悲しい話題など，文章の内容に応じてキャラクタアニメーションを動的に変化させる必要があると考えられる．以上のような点を改善し，詳細内容表示画面をより理解しやすいものとすることが重要である．

評価実験について

今回の実験は，1人30分程度の短い実験であり，最初の印象として，Bubble Browserが子供にとって楽しい結果となったが，“飽きないかどうか”についての評価が十分にはできていない．そこで，実際にBubble Browserを長期間使用してもらい，他の一般的なWebブラウザと比較して，飽きずに使用し続けてもらえるかどうかを検証する必要がある．現状では，Bubble Browserは，どのようなWebページに対しても，常に同じような海中アニメーション画面と絵本による詳細内容表示画面を提示するため，長期間の実験では，今

回行った実験よりも，Bubble Browser を楽しいと評価する被験者は減ると考えられる．そのため，Web ページの内容によって使用するアニメーションやキャラクタ，台詞の種類を変化させることで，長期間使用する際に飽きることを防止するための機能の検討も必要である．

2.4 むすび

本章では，難解なページであっても子供が興味を持って分かりやすく閲覧できる Web ページの提示を目的として，泡メタファを用いた子供向け Web ブラウザの設計と実装を行った．提案したブラウザでは，子供の特徴を考慮したデザインとして泡メタファを用いた海中アニメーションによる Web ページの提示を行う．具体的には，Web ページをコンポーネントに分割し，コンポーネントごとに泡を生成して表示する．コンポーネントの画像情報，テキスト情報を泡上に表示し，さらに，コンポーネントの特徴を泡のサイズ，色や上昇スピードとして視覚的に伝える．また，コンポーネントの詳細内容は子供が楽しく理解しやすいように絵本形式へ変換して提示する．これにより，子供にとって理解が難しい一般向け Web ページでも，楽しく分かりやすく Web 閲覧することができる．

提案ブラウザの有効性を確認するため，13 名の子供による評価実験を行った．その結果，比較対象として用いた漢字を平仮名に変換して提示するよりブラウザよりもページ内容の質問に対する正答率が 20%向上したことから，子供にとってわかりやすい提示を実現できたと考える．また，Web ページを元の形態のまま提示するブラウザよりも楽しいと答えた被験者が 50%増加したことから，一般のブラウザよりも画像を用いることで子供の興味をひく提示を実現できたと考える．

第3章

Web検索結果の子供向けリランク手法

3.1 まえがき

第2章ではWebページ自体を子供にとって興味をひきやすい形式に変換することで，子供のWeb閲覧の支援を行った．しかし，変換せずとも，元々子供にとって十分興味をひき，分かりやすいようなページも存在する．そのようなページが存在する場合は，変換せずに，そのままページを提示するだけで十分であると考えられる．しかし，そのようなページは一般向けページほど数が多くないため，検索エンジンの検索結果から探し出すのが困難である．本章では，Web検索を行う際に，そのような子供向けのWebページを容易に取得可能とするための検索機構について検討する．

Web検索を行う場合，ランキング形式で表示される検索結果の中から，自分が必要としている情報を選択する必要がある．しかし，一般的な検索エンジンはもちろん，子供向けの検索エンジンでも，基本的に検索結果のランキングは子供向けでない．例えば，一般的な検索結果において，Wikipediaの記事が上位であることが多いが，このようなページは文章が多く，難解な表現が多用されているため，子供が必要としている分かりやすいページとはいえない．また，子供向け検索エンジンにおいても，あらかじめ登録されたお勧めサイト以外のランキングは一般的なものと同じであり，お勧めサイトであっても必ずしも子供向けであるとはいえない．そのため，現状の検索エンジンでは，子供が必要とする情報がランキングの上位であるとは限らない．子供には，検索結果から自分が求めている結果を選択するのが苦手であり，検索結果の上位5件程度しか閲覧しないという特徴があるため [16]，検索結果のランキングでは，子供向けのページを上位にランクする必要がある．

そこで本章では，子供向けページを上位にランクすることを目的とし，一般的な検索エンジンの検索結果を子供向けにリランクする手法を提案する．提案手法では，1.2.1項で述べたような子供のWebページに対する嗜好の調査に関する既存研究で得られた知見 [58,59,73,74] に基づき，子供にとって，見た目が見やすく，役立ち，内容が分かりやすいようなページを子供の興味をひきやすい子供向けページと定義し，一般向けページと子供向けページが混在したページ集合から子供向けページを判定できるよう，ページの子供向け度合いをスコア化する．具体的には，ページ中の画像やアニメーションの量，文章の量，色の数などの構成に関する指標，ページ中の文章の平均文字数，漢字やアルファベットの量，子供向け表現の数などの文章に関する指標を設定する．そして，各指標のスコアを組み合わせて各ページの子供向けスコアを決定し，検索エンジンの検索結果を子供向けスコアの降順に並び変えることで，子供向けのリランクを行う．これにより，子供向けのページが検索結果上位にランクされるため，子供はランキング下位まで探す手間なしに，子供向けのページを容易に得ることできる．

以下では，3.2節で提案手法である子供向けリランク手法について述べる．その後，3.3節で評価実験について述べ，最後に3.4節で本章のまとめを行う．

3.2 子供向けリランク手法

3.2.1 子供向けページの定義

本研究では子供向けページとして，1.3.1項で述べた子供向け情報に合致するものを上位にランクする．具体的には，Webページの構成と内容について，網羅的に以下の特徴量を考慮する．

- Webページの構成
  - 子供の興味をひくように，画像やアニメーションが使用されている [58,59,73,74].
  - 子供にとって分かりやすいように，テキストは適度な量である [58,59,73,74].
  - リンクはどこをクリックすべきか分かるように把握しやすい量である [74].
  - 子供にとってスクロール操作は負担であるため，スクロール操作が少量で済むページサイズである [9,74].

- 子供の興味をひくように，カラフルな色使いである [73,74].

- Webページの文章

  - 子供の興味を削がないように，一般的な文と比べると，長さは短めで，一文に含まれる情報が少ない [89].
  - 子供にとって分かりやすいように，漢字，アルファベットや難解な表現が少ない [58,59,89].
  - 子供にとって親しみやすいように，語りかけの表現などの子供向け特有の表現が含まれている [74,89].
  - 子供にとって分かりやすいように，難解な表現が使用されず，理解が容易な文である [58,59,73,74,89].

3.2.2 指標の設計

3.2.1 項で述べた定義に基づき，一般向けページと子供向けページが混在するページ集合から子供向けページを判定できるよう，表 3.1 に示す指標を設定した．これらの指標は，既存研究で得られた知見を基に，試験的に設計しており，3.3 節の評価実験で，各指標の効果を調査し，最終的に組み合わせるべき指標について考察する．設定した指標は，構成に関する指標 6 個，文章に関する指標 5 個の計 11 個であり，これらの指標を用いて Web ページのスコア化を行い，スコアが大きいほど子供向けとする．

以降では，それぞれの指標の詳細を述べる．指標の値域については，値が小さいほど子供向けである指標は $-1 \sim 0$，値が大きいほど子供向けである指標は $0 \sim 1$ の範囲となるよう，正規化を行う．これは，各指標を組み合わせてページのスコアを決定する際に，値が大きいほど良い指標は加算，値が小さいほど良い指標は減算して組み合わせるためである．

構成に関する指標

子供向けのページとして，Web ページのサイズやコンテンツ量というページの構成は，子供にとって情報を把握しやすく，興味を持てる形式である必要がある．そこで，以下の 6 つの指標を定義する．

表 3.1: スコア化に用いる指標

タイプ	指標	説明
構成	*Size*	ページの面積
	ImageRate	ページ中の画像やアニメーションの割合
	TextRate	ページ中の文章の割合
	AncorRate	リンクの数
	Component	コンポーネントの数
	Color	使用されている色の数
文章	*KanjiRate*	テキスト中の漢字，アルファベットの割合
	ChildrenExpression	テキスト中の子供向け表現の割合
	DifficultExpression	テキスト中の難解表現の割合
	Easy	テキストの難易度
	SentenceLength	テキストの平均文字数

- ページのサイズ (*Size*)

  子供はスクロールを苦手とする特徴があるため，スクロールが必要な大きいサイズのページは子供向きのページとはいえない．そこで，ページの面積に基づいた指標を *Size* とし，図 3.1 (a) に示すような面積が小さければ子供向けページであると判断する．具体的には，ページ i の $Size_i$ は，式 (3.1) に基づき，計算する．値が小さいほど子供向けなので，値域は $-1 \leq Size \leq 0$ とする．

$$Size_i = \begin{cases} -1 & (size\ of\ page_i \geq maxsize) \\ -\dfrac{size\ of\ page_i}{maxsize} & (size\ of\ page_i < maxsize) \end{cases} \tag{3.1}$$

  ここで，$size\,of\,page_i$ はページ i の面積，$maxsize$ はページの面積の最大値であり，ランダムな 500 ページの面積を調査した結果に基づき，$1,000 \times 5,000$ $[pix]$ と設定した．

- 画像量 (*ImageRate*)

  画像やアニメーションは，子供の興味をひき，理解のサポートにもなるため，子供向けページに不可欠な要素であると考えられる．そこで，画像やアニメーションの量を判断する指標を *ImageRate* とし，図 3.1 (b) に示すようなページ全体の面積のうちの

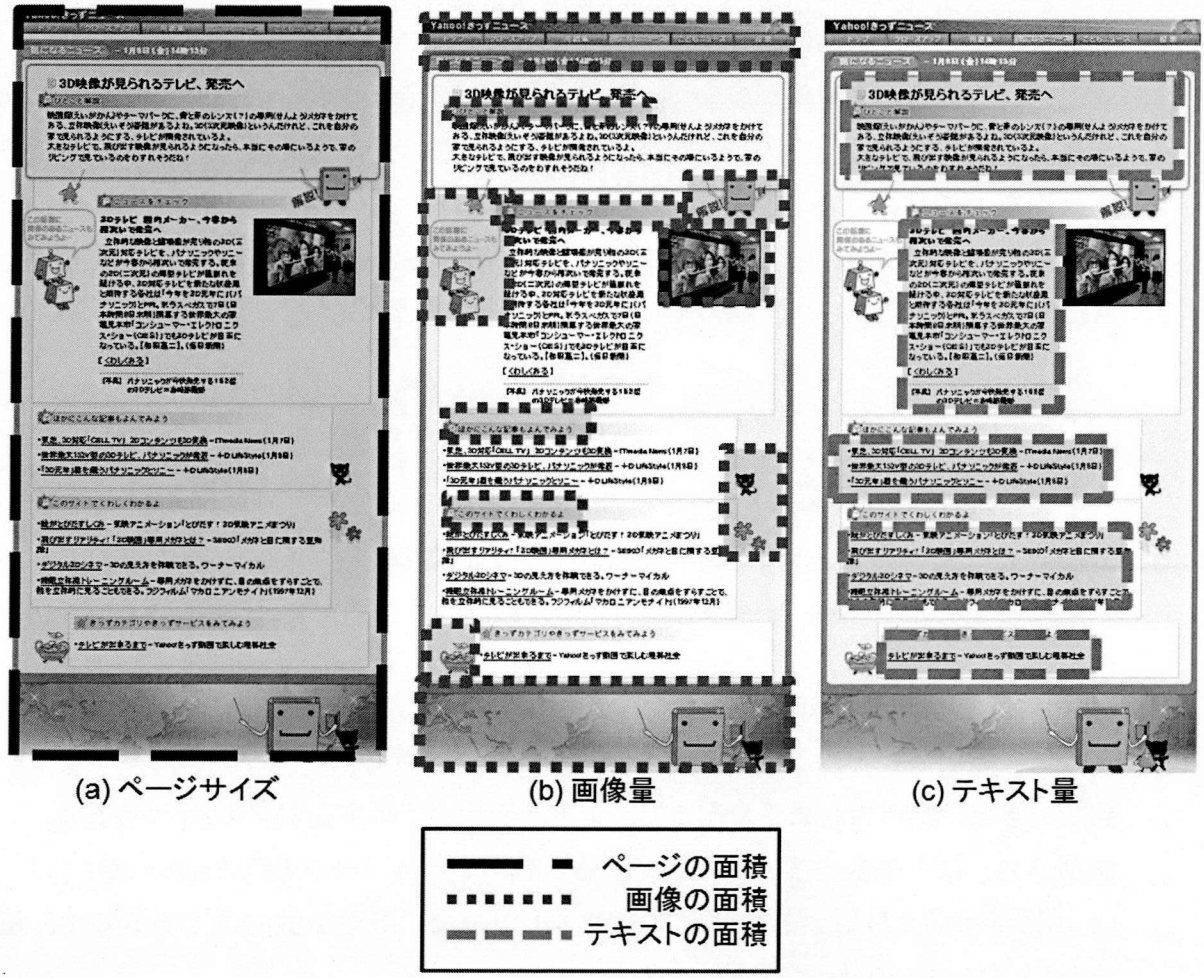


図 3.1: 指標の例 ($Size$, $ImageRate$, $TextRate$)

画像やアニメーションの占める面積の割合が大きいほど，子供向けのページであると判断する．具体的には，ページ i の $ImageRate_i$ は式 (3.2) に基づき計算する．値が大きいほど子供向けであるので，値域は $0 \leq ImageRate \leq 1$ とする．

$$ImageRate_i = \frac{\sum_{j=1}^{N} size\ of\ image_j^i}{size\ of\ page_i} \tag{3.2}$$

ここで，$size\ of\ image_j^i$ はページ i 内の j 番目の画像とアニメーションの面積，N はページ i に含まれる画像とアニメーションの数，$Size_i$ は前述したページ i の面積である．画像とアニメーションの面積は，HTML テキストより，$\langle img \rangle$，$\langle script \rangle$ などの HTML タグ内の情報を抽出して決定する．

- テキスト量 ($TextRate$)

  文章量が多いと，子供にとっては面白みに欠け，ページを閲覧する意欲を持てないと考えられるため，文章量が多いページは子供向けであるとはいえない．そこで，文章量を判断する指標を $TextRate$ とし，図 3.1 (c) に示すようなページ全体の面積のうちの文章の面積の占める割合が小さいほど子供向けであると判断する．具体的には，ページ$_i$ の $TextRate_i$ は，式 (3.3) に基づき計算する．値が小さいほど子供向けであるので，値域は $-1 \leq TextRate \leq 0$ とする．

$$TextRate_i = -\frac{\sum_{j=1}^{N} length\ of\ text_j^i \cdot fontsize}{size\ of\ page_i} \tag{3.3}$$

  ここで，$length\ of\ text_j^i$ はページ i の j 番目の文の文字数，N はページ i の文の数，$fontsize$ は文字のフォントサイズであり，今回は標準の 16 [point] を全ての文字のサイズと設定した．$Size_i$ は前述したページ i の面積である．

- リンク量 ($AnchorRate$)

  リンクの量が多いと，子供はどのリンクを辿れば閲覧したいページに辿りつけるのか判断できなくなってしまう．そこで，リンクの量を判断する指標を $AnchorRate$ とし，値が小さいほど子供向けのページであると判断する．ページ i の $AnchorRate_i$ は，式 (3.4) に基づいて計算する．リンクテキストが少ないほど子供向けであるので，値域は $-1 \leq AnchorRate \leq 0$ として定義する．

$$AnchorRate_i = -\frac{number\ of\ anchor_i}{maxnumber\ of\ anchor} \tag{3.4}$$

  ここで，$number\ of\ anchor_i$ はページ i に出現するリンクの数，$maxnumber\ of\ anchor$ はリンクの数の最大値であり，ランダムな 500 ページを調査した結果に基づき，300 個と設定した．

- コンポーネント ($Component$)

  コンポーネントとは，Web ページ内の関連する情報が集まったブロックである．コンポーネントの数が多いほど，ページの構成は複雑になる傾向があり，子供にとっては，注目すべきコンテンツを判断するのが難しくなる．そこで，コンポーネントの数に関する指標を $Component$ とし，コンポーネントの数が少ないほど，子供向けのページであると判断する．具体的には，筆者の所属する研究グループが提案した方

式 [62] でコンポーネントを抽出し，ページ i の $Component_i$ は式 (3.5) に基づき，計算する．数が少ないほど子供向けであるので，値域は $-1 \leq Component \leq 0$ とする．

$$Component_i = \begin{cases} -1 & (number\ of\ comp_i \geq maxnum) \\ -\dfrac{number\ of\ comp_i}{maxnum} & (number\ of\ comp_i < maxnum) \end{cases} \tag{3.5}$$

ここで，$number\ of\ comp_i$ はページ i のコンポーネントの数，$max\ of\ comp$ はコンポーネント数の最大値であり，ランダムな 500 ページを調査した結果に基づき，20 個と設定した．

- 色数 ($Color$)
  使用されている色の数が少ないページよりは，色の数が多く，カラフルな見た目であるページの方が，子供の興味をひく表示と考えられる [73,74]．そこで，ページに表示されている色の数を判断する指標を $Color$ とし，色の数が多いほど子供向けページと判断する．ページ i の $Color_i$ は式 (3.6) に基づき，計算する．値が大きいほど子供向けであるので，値域は $0 \leq Color \leq 1$ とする．

$$Color_i = \frac{number\ of\ color_i}{maxnumber\ of\ color} \tag{3.6}$$

  ここで，$number\ of\ color_i$ はページ i で使用されている色の数で，ページで使用される色については，JPEG 形式の Web ページのスクリーンショット画像から異なる色の数をカウントし，$count\ of\ color_i$ を決定する．$maxnumber\ of\ color$ は色の数の最大値であり，JPEG 画像の最大色数である 1670 万色と設定した．

文章に関する指標

子供向けのページとして，Web ページの文章は，子供にとって分かりやすく，興味を持てる形式である必要がある．そこで，以下の 5 つの指標を定義する．

- 漢字・アルファベット量 ($KanjiRate$)
  漢字，アルファベットの量が多いほど，その文章は子供にとって理解するのが困難であると考えられる．そこで，文章中の漢字，アルファベットの量に関する指標を $KanjiRate$ とし，ページに含まれる文章中の漢字とアルファベットの文字数の占める割合が少ないほど，子供向けのページであると判断する．ページ i の $KanjiRate_i$

は式(3.7)に基づき，計算する．値が小さいほど子供向けであるので，値域は $-1 \leq KanjiRate \leq 0$ とする．

$$KanjiRate_i = -\frac{number\ of\ kanji_i + number\ of\ alphabet_i}{length\ of\ text_i} \tag{3.7}$$

ここで，$numberofkanji_i$ はページ i の文章に含まれる漢字の文字数の合計，$numberof alphabet_i$ はページ i の文章に含まれるアルファベットの文字数の合計，$lengthoftext_i$ はページ i に含まれる文章の文字数の合計である．

- 子供向け表現の量 ($ChildrenExpression$)

  “～しているよ” などの子供向け表現が多く含まれているページは，子供にとって親しみやすく，内容も容易な文章で書かれている可能性が高い．そこで，文章に含まれる子供向け表現に関する指標を $ChildrenExpression$ とし，子供向け表現が多く含まれているほど，子供向けページと判断する．子供向け表現は，Yahoo! きっず，キッズgoo，NHK週刊こどもニュース[1] を始めとする子供向けページ1,000ページから形態素，文末表現などの頻出表現を抽出して利用する．ページ i の $ChildrenExpression_i$ は式(3.8)に基づき，計算する．値が大きいほど子供向けであるので，値域は $0 \leq ChildrenExpression \leq 1$ とする．

$$ChildrenExpression_i = \frac{\sum_{j=1}^{N} number\ of\ children\ expression_j}{number\ of\ term_i} \tag{3.8}$$

  ここで，$number\ of\ children\ expression_j$ はページ i の文章の j 番目に出現する子供向け表現，N はページ i の文章中の子供向け表現の出現数であり，あらかじめ作成した辞書の子供向け表現と合致する数とする．$number\ of\ term_i$ はページ i の文章に含まれる形態素，文末表現の総数である．

- 難解表現の量 ($DifficultExpression$)

  難解な表現が多く含まれるページでは，子供は読む意欲を失い，内容を理解することが困難であると考えられる．そこで，文章に含まれる難解な表現に関する指標を $Difficult\ Expression$ とし，難解な表現の量が少ないほど，子供向けページであると判断する．難解表現は，Yahoo! ニュース1,000ページ，Wikipediaの記事1,000ページから頻出表現を抽出して利用する．子供にとって一般向けに作成されたニュー

[1] <NHK週刊こどもニュース: http://www.nhk.or.jp/kdns/>

スページ，また専門用語などを解説するようなページは難解であると考えられるため，Yahoo! ニュースと Wikipedia を用いた．ページ i の $DifficultExpression_i$ は式 (3.9) に基づき，計算する．値が小さいほど子供向けであるので，値域は $-1 \leq DifficultExpression \leq 0$ とする．

$$DifficultExpression_i = -\frac{\sum_{j=1}^{N} number\ of\ difficult\ expression_j}{number\ of\ term_i} \tag{3.9}$$

ここで，$number\ of\ difficult\ expression_j$ はページ i の文章の j 番目に出現する難解表現，N はページ i の文章中の難解表現の出現数であり，あらかじめ作成した辞書の難解表現と合致する数とする．$number\ of\ term_i$ はページ i の文章に含まれる形態素，文末表現の総数である．

- 難易度 ($Easy$)

  単純に漢字・アルファベットの量，難解な表現の量という部分的な要素だけでなく，文章全体の難易度が低い方が，子供にとって理解が容易なため，文章全体の難易度に関する指標を $Easy$ とし，難易度が低いほど子供向けのページと判断する．難易度は，教科書から作成したコーパスを用いて日本語の文章の難易度を推定するツールである帯 [81] を利用する．帯によって推定される難易度を用い，ページ i の $Easy_i$ は式 (3.10) に基づき，正規化する．値が小さいほど子供向けであるので，値域は $-1 \leq Easy \leq 0$ とする．

$$Easy_i = -\frac{level_i}{13} \tag{3.10}$$

  ここで，$level_i$ は帯で推定したページ i のテキストの難易度であり，13 は帯で推定する難易度の最大値である．

- 文の平均文字数 ($SentenceLength$)

  一文が長いほど，一文に含まれる情報が多くなり，子供がその文の内容を理解することが難しくなると考えられる．ここで，一文の長さとは，漢字を平仮名に変換してカウントしたときの文字数と定義する．文字数の少ない文は，含まれている情報が簡潔で，子供にとって容易に理解できる文と考えられる．そこで，一文の長さに関する指標を $SentenceLength$ とし，一文の文字数が少ないほど子供向けのページと判断する．具体的には，ページに含まれる全ての文の文字数を平均し，$SentenceLength$ と

し，ページ i の $SentenceLength_i$ は式(3.11)に基づき計算する．値が小さいほど子供向けであるので，値域は $-1 \leq SentenceLength \leq 0$ とする．

$$SentenceLength_i = \begin{cases} -1 & (average\,len_i \geq maxlen) \\ -\dfrac{average\,len_i}{maxlen} & (average\,len_i < maxlen) \end{cases} \tag{3.11}$$

ここで，$average\,len_i$ はページ i の文章中の全ての文の平均文字数である．$maxlen$ は一文の文字数の最大値であり，ランダムな500ページの一文の平均文字数を調査した結果に基づき，100文字と設定した．

3.2.3 リランク手順

3.2.2項で述べた指標に基づき，Webページのスコアを算出し，リランクを行う．まず，以下のような手順でWebページの子供向けスコアを計算する．

(1) クロールを行い，Webページを収集する．

(2) 子供が安全にWeb検索を行えるよう，フィルタリングで有害なページを削除する．

(3) 指標に基づき，各ページのスコアを算出する．具体的には，$-1 \sim 1$ の各指標のスコアを加算して組み合わせ，各ページのスコアとする．

そして，ユーザがクエリを発行すると，以下の手順でリランクを行う．

(1) ユーザがクエリを入力する．

(2) クエリに対する検索結果を検索エンジンより取得する．

(3) 取得した検索結果のWebページのスコアを取得する．

(4) スコアの降順にリランクして表示する．

3.3 評価実験

本節では，各指標の有効性，指標間の関係を調査し，その結果に基づいて指標を組み合わせた際の有効性を検証するために行った評価実験について述べる．

3.3.1 データセット

提案手法を評価するためには，実際に子供にWebページを評価してもらった正解データが必要である．そこで，Webページとそのページに対する子供による評価点の集合をデータセットと定義し，以下のようにデータセットを作成した．まず，Yahoo!きっず，キッズgooにおける週間検索キーワードランキングの2009年12月から2010年3月のランキングの上位（10位まで）より選んだ“地球温暖化”，“介助犬”，“ゲーム”，“うらない”，“冬至”，“百人一首”の6個のキーワードを実験に使用するクエリとし，各クエリごとにYahoo! JAPAN，Yahoo!きっずの検索結果のランキングのそれぞれの上位から25のWebページを取得した．そして，合計300のWebページを小学校低学年(1, 2年生)から高学年(5, 6年生)の男女33名に評価してもらい，Yahoo! JAPAN用，Yahoo!きっず用のデータセットとした．被験者の内訳は，低学年4人（男子2人，女子2人），中学年10人（男子6人，女子4人），高学年19人（男子15人，女子4人）であった．

具体的には，それぞれの被験者に，Webページを30秒程度閲覧してもらい，“読みたいと思うか？”，“見た目が見やすいか？”，“勉強に役立つか？”，“内容が分かりやすいか？”の4つの質問にYes，Noの2択で答えてもらい，1ページ当たり4人から7人による評価を得た．そして，Yesを1点，Noを0点とし，各ページごとに点数を平均したものをそのページの子供向けスコアとし，子供向けスコアの降順に並び変えたものを理想ランキングとし，データセットを作成した．この際，スコアが同じページは，Yahoo! JAPANのランキングに基づいた順位とした．4つの質問で評価を行ったのは，Webページを子供向けであると判断するためには，ページの構成や見た目といった外観に基づく指標，文章など内容に基づく指標など多面的な評価が必要であり，それらを個々に検証するためである．具体的には，“読みたいと思うか？”の質問で子供にとって興味を持ちやすいページがどのようなものなのか，“見た目が見やすいか？”の質問で子供にとって見やすいと感じるページがどのようなものなのか，“勉強に役立つか？”の質問で子供にとって学習になるページがどのようなものなのか，“内容が分かりやすいか？”の質問で子供にとって理解が容易なページがどのようなものなのかを調査する．

ここで，Yahoo! JAPAN，Yahoo!きっずのそれぞれでデータセットを作成したのは，Yahoo! JAPANとYahoo!きっずでは検索結果に含まれるページが大きく異なるためである．Yahoo!きっずのランキングでは，クエリに対応する登録サイトがあれば，そのサイトを検索結果上位に表示した後，Yahoo! JAPANと同様のランキングを表示するが，Yahoo!きっ

ずに登録されていないページは表示されない．そのため，Yahoo!きっずのランキングによるデータセットでは，Wikipediaなどの明らかに子供向けでないページは含まれておらず，Yahoo! JAPANのランキングと比べて，子供向けページが多く含まれる傾向がある．

3.3.2 評価指標

評価指標には，正解データがクエリへの多段適合度を持つ場合の順位付き検索結果の性能を測るNDCG (Normalized Discounted Comulative Gain) [55] を用いた．クエリ q に対するNDCGは，適合度の高い順に並べた理想的な結果とのずれを表す指標で，以下の式より求める．

$$NDCG_q = \frac{1}{IDCG_q}\left(rel_1 + \sum_{i=2}^{l}\frac{rel_i}{log_2 i}\right) \tag{3.12}$$

rel_i は検索結果 i 番目のページのデータセットで定義されたスコア，l は検索数であり，子供が一般的に閲覧すると言われている上位5件で評価を行った．$IDCG_q$ はクエリ q に対する $NDCG_q$ の理想値，つまりデータセットで定義された理想ランキングの $NDCG$ 値である．

3.3.3 実験結果

提案手法によるリランク結果とYahoo! JAPAN，Yahoo!きっずによるランキング結果を比較した結果について述べる．

リランクを行う際には，教師付き機械学習を用いて，3.2.2項で述べた指標を適切に組み合わせることが有効であると考えられる．一般的に，事例とそれを表す複数の指標からなる高次元ベクトルを用いて学習を行う際，指標の次元が多い場合は計算コストが高くなりやすく，さらに，正解の推定に関してあまり意味を持たない不要な指標により，最良の推定を行うことができないことがある [44,75]．具体的には，正解の推定に関して無関係な指標は，データにノイズを加え，精度に影響を与えてしまう．さらに，相関した指標のグループが含まれる場合は，これらの指標は，実際には同一の特徴を表している場合があり，こうした指標が多数存在していると，精度が影響を受ける可能性がある．そのような場合，指標単独での効果や指標間の相関性を考慮して，ノイズや相関の影響を最小限に抑える最適な指標を選択することで精度を向上させることができる [44]．そこで，高精度なリランクを行うために，まず，単独指標でリランクした際の精度の結果を調査し，さらに，指標間の相関性を考慮し，最終的に組み合わせる指標を決定する．

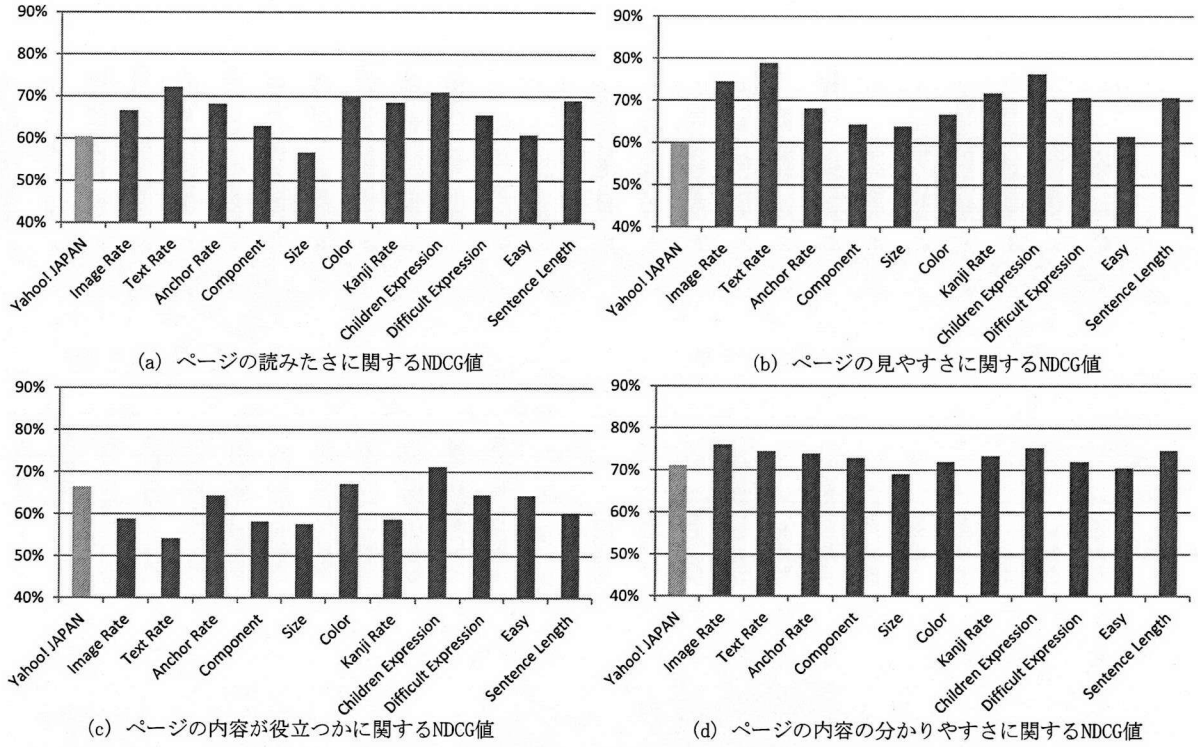


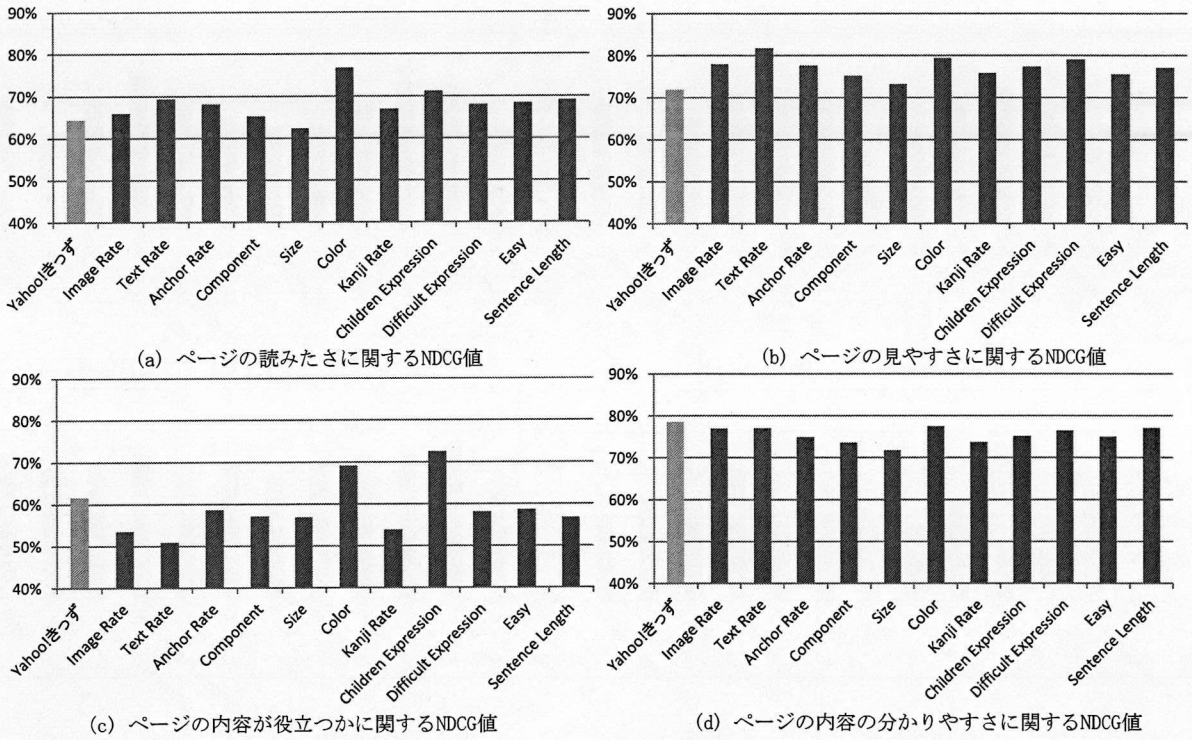
図 3.2: 各指標の NDCG の平均値 (Yahoo! JAPAN)

ここで，Yahoo!きっずのデータセットは子供向けページを多く含むため，Yahoo! JAPAN よりも NDCG 値は高くなる傾向がある．具体的には，Yahoo!きっずの平均 NDCG 値が 0.68 であったのに対して，Yahoo! JAPAN の平均 NDCG 値は 0.60 となった．データセットでは，“読みたいか？”，“見た目が見やすいか？”，“勉強に役立つか？”，“内容が分かりやすいか？” の 4 つの質問で理想ランキングを定義しているため，それぞれの質問ごとに結果を述べる．

単独指標によるリランク

Yahoo! JAPAN のデータセットにおいて指標を単独に用いたリランクの NDCG の平均値を図 3.2 に，Yahoo!きっずのデータセットにおいて指標を単独に用いたリランクの NDCG の平均値を図 3.3 に示す．グラフでは，緑が Yahoo! JAPAN，Yahoo!きっずのランキング，青が構成に関する指標，赤が文章に関する指標によるリランクの結果を示す．

図 3.2 (a)，図 3.3 (a) に示すページの読みたさについては，Yahoo! JAPAN，Yahoo!きっ



(a) ページの読みたさに関するNDCG値

(b) ページの見やすさに関するNDCG値

(c) ページの内容が役立つかに関するNDCG値

(d) ページの内容の分かりやすさに関するNDCG値

図 3.3: 各指標の NDCG の平均値 (Yahoo!きっず)

ずともに，*Size* のみ NDCG 値が 2%下がり，それ以外の指標では NDCG 値が 1%から 14%上がった．*Size* の NDCG 値が下がったのは，必ずしもサイズが小さい程読みたいと感じるわけではなく，最適値が存在し，それより小さいページは興味を失う傾向があるためと考えられる．ただし，Wikipedia の記事のようにスクリーンの数倍以上のサイズのページは読みたくないと判断した子供は多かったので，そのようなページを排除する補助的な指標として使用することが有効と考える．精度の上がった指標の中で，特に *TextRate*，*Color*，*ChildrenExpression* は NDCG 値が 7%から 18%向上しており，子供にとっては，テキストが少なく，カラフルであり，さらに子供向け表現が含まれるページを読みたいと思うものと考えられる．

図 3.2 (b)，図 3.3 (b) に示すページの見やすさについては，Yahoo! JAPAN では全ての指標で 2%から 20%，Yahoo!きっずでも全ての指標で 4%から 10%ほど NDCG 値が向上していた．特に，*ImageRate*，*TextRate* は，Yahoo! JAPAN，Yahoo!きっずどちらにおいても単独に指標を用いるだけで，5%から 18%も NDCG 値が向上しており，子供にとっては，画像量が多く，テキスト量が少ないページならば見た目に見やすいと思う傾向が強いもの

と考えられる．

図3.2 (c)，図3.3 (c) に示すページが役立つかどうかについては，*ChildrenExpression*, *Color* はNDCG値が1%から11%向上し，それ以外の指標ではNDCG値は下がっていた．これは，30秒程度ページを閲覧しただけでは勉強に役立つかどうかを判定するのが難しく，文章が多いページを単純に勉強になると判断した子供が多かったため，多くの指標が有効に働かなかったものと考えられる．*ChildrenExpression* は，あらかじめ定義した辞書に“なぜ”，“学ぼう”などの子供用の勉強ページに含まれるような表現が多数含まれていたため，精度良く判定できたものと考えられる．*Color* についても，子供用にカラフルに作成された勉強用のページを判定できたものと考えられる．

図3.2 (d)，図3.3 (d) に示すページの内容の分かりやすさについては，Yahoo! JAPANではどの指標でもNDCG値は5%以下の向上にとどまっており，Yahoo!きっずでは全ての指標で精度が下がった．これは，画像，テキストの両方をある程度用いて説明しているページを分かりやすいと見なす子供が多く，単純にテキストが少ないほど分かりやすいと判断することが不十分であったと考えられる．特に，Yahoo!きっずでは，多くのページが小学校高学年程度の子供にとって十分に分かりやすいため，むしろ，文章が多めで詳しく解説されているページを分かりやすいと見なす傾向があった．そのため，学年が上がるほど文章は多めが良いというように，各指標のスコアを学年ごとで適切な値とする必要がある．

これらの結果より，指標を単独に用いただけでも，多くの場合でYahoo! JAPAN，また元々精度の高いYahoo!きっずのランキングの精度を上回ることができることを確認した．また，どのような子供向けページを上位にランクするかで効果の大きい指標が異なることが分かる．“読みたさ”，“見やすさ”，“分かりやすさ”については画像量やテキスト量などのページの構成が子供向けであることが共通して必要であり，“役立つか”についてはそのような構成とはあまり関係がない．そして，“読みたさ”，“役立つか”，“分かりやすさ”については文章が子供向けであることが共通して重要である．しかし，“役立つか”，“分かりやすさ”については文章が容易なだけではなく，年齢に応じた適度なテキスト量，難易度である必要がある．つまり，“読みたさ”，“見やすさ”は組み合わせて使うことができるが，“分かりやすさ”，“役立つか”は単独で使うべきであると考える．

指標間の関係

指標単独でのリランクの結果を示したが，さらに精度を向上させるためには，指標を組み合わせることが有効と考えられる．そこで，指標間の相関性を考慮し，組み合わせる指標を検討する．3.2.1 項で述べた 11 の指標のうち，以下に示す指標は独立性が低く，互いに相関性が強いと考えられる．

- *Size*（ページのサイズ），*Component*（コンポーネント数）：
  ページのサイズが大きいほど，含まれるコンポーネント数は増え，ページの構成が複雑になる傾向がある．そのため，ページのサイズとコンポーネント数は，ページの構成の複雑さを示すために用いる類似した指標であり，相関性が強いと考えられる．

- *ImageRate*（画像量），*Color*（色の数）：
  ページ内に多くの画像が用いられるほど，色の数が増え，見た目に興味をひきやすい構成となる傾向がある．そのため，画像量と色の数は，どちらも見た目に興味をひく構成に関係する要素であり，相関性が強いと考えられる．

- *KanjiRate*（漢字・アルファベット量），*ChildrenExpression*（子供向け表現の数），*DifficultExpression*（難解表現の数），*Easy*（難易度），*SentenceLength*（平均文字数）：
  文章に関する指標はどれも子供にとって分かりやすい文章であるかどうかを推定する指標であり，独立性が低いと考えられる．具体的には，子供にとって分かりやすく難易度の低い文章は，漢字やアルファベット，難解な表現が少ない一方，子供向け表現が多く，平均文字数が短い傾向があると考えられる．

そこで，これらの指標間の相関性を調査するために，指標ペア間の偏相関係数を算出した．偏相関係数は，複数の変数がある場合に，他の変数の影響を取り除いて，2 つの変数の相関関係を調査するために使用される．表 3.2 に独立でないと考えられる指標間の偏相関係数を示す．偏相関係数は 3.2 節で述べた式により算出した各指標のスコアを用いて算出した．一般的に，相関係数は ± 0.2〜0.4 で低い相関があるとされる [43].

Size と *Component* については，偏相関係数は 0.654 と他と比較しても大きく，相関性が強いことが分かる．*ImageRate* と *Color* については，偏相関係数は 0.283 であり，*Size* と *Component* 間ほどではないが，低い相関が認められる．*KanjiRate*, *ChildrenExpression*,

表 3.2: 指標間の偏相関係数

指標のタイプ	指標のペア	相関係数
構成	*Component* と *Size*	0.654
	ImageRate と *Color*	0.283
文章	*ChildrenExpression* と *DifficultExpression*	0.312
	KanjiRate と *Easy*	0.244
	KajiRate と *ChildrenExpression*	0.200
	Easy と *SetenceLength*	0.170
	DifficultExpression と *Easy*	0.080
	ChildrenExpression と *SentenceLength*	0.062
	KanjiRate と *DifficultExpression*	−0.001
	ChildrenExpression と *Easy*	−0.020
	KanjiRate と *SentenceLength*	−0.086
	DifficultExpression と *SentenceLength*	−0.094

Difficult Expression, *Easy*, *Sentence Length* については，10 の指標ペアのうち，*Kanji Rate* と *ChildrenExpression*，*Kanji Rate* と *Easy*，*Children Expression* と *Difficult Expression* の 3 ペアで 0.2 以上の偏相関係数となり，*Easy* と *SentenceLength* でも 0.170 と 0.2 に近い偏相関係数となった．つまり，漢字・アルファベットの量と子供向け表現の数，難易度や子供向け表現の数と難解表現の数，難易度と平均文字数は，文章の指標の中でも，互いに低い相関があり，独立性がそれほど高くないことが分かる．

上記以外のペアについては，相関係数は 0.1 以下となった．これは，*SentenceLength* が，短いキーワードで表されたリンク文字列であるメニューや広告などの影響を受けやすく，そのようなページでは平均文字数が短くなる可能性があることが一つの原因であると考えられる．つまり，*SentenceLength* については，他の文章に関する指標とは，独立性が比較的高いと考えられる．

Children Expression や *Difficult Expression* が，他の指標と異なり，あらかじめ用意した辞書内の表現に依存していることも一つの原因と考えられる．*Kanji Rate* と *Children Expression* については，辞書内に定義した子供向け表現には，直接的に平仮名が多い傾向

があるため，相関が見られたが，それ以外の指標との間には大きな相関は見られなかった．

このように，11の指標の中には，独立性が十分でなく，実際に他の指標との相関の高い指標が含まれていることが分かる．そのため，指標を組み合わせる際には，互いに相関性の強いグループから代表的な指標を選択して用いることが有効と考えられる．

指標の組合せによるリランク

前述の議論に基づいて，指標を組み合わせた際の有効性を検証する．具体的には，教師付き機械学習を用いたランキング関数の学習の代表的な手法である Ranking SVM [36] を用いて，指標を組み合わせて利用した場合の性能を評価した．実装には svm-rank[2] を用い，カーネルには線形カーネルを用いた．Ranking SVM による学習を行うことで，各指標を適切に組み合わせることが可能となる．ここで，学習は，5種類のクエリに対応する250ページで行い，残り1種類のクエリに対応する50ページで評価を行うことを1セットとし，6種類のクエリ全てで評価できるように，6セットの評価を行った．

指標を組み合わせた際の，Yahoo! JAPAN のデータセットにおけるリランクの NDCG 値の平均を図3.4(a) に，Yahoo!きっずのデータセットにおけるリランクの NDCG 値の平均を図3.4(b) に示す．緑色が Yahoo! JAPAN，Yahoo!きっずのベースラインとなる NDCG 値，赤色の *All* が単純に全ての指標を用いた場合の NDCG 値，青色の *Cut* が，指標単独によるリランク，指標間の関係の結果に基づいて独立性の観点から不要と考えられる指標を取り除いた場合の NDCG 値を示す．具体的に *Cut* は，互いに相関性が強いと考えられるグループの中で，指標単独での NDCG 値の低い指標を除いたものであり，以下のように決定した．

- *Component*, *Size* :
  NDCG 値の高い *Component* を使用．

- *ImageRate*, *Color* :
  NDCG 値の高い *Color* を使用．

- *KanjiRate*, *ChildrenExpression*, *DifficultExpression*, *Easy*, *SentenceLength*:
  NDCG 値の高い *ChildrenExpression*，および，ある程度 NDCG 値が高く他の指標と相関の小さい *SentenceLength* を使用．

[2] <SVM-rank: http://www.cs.cornell.edu/People/tj/svm_light/svm_rank.html>

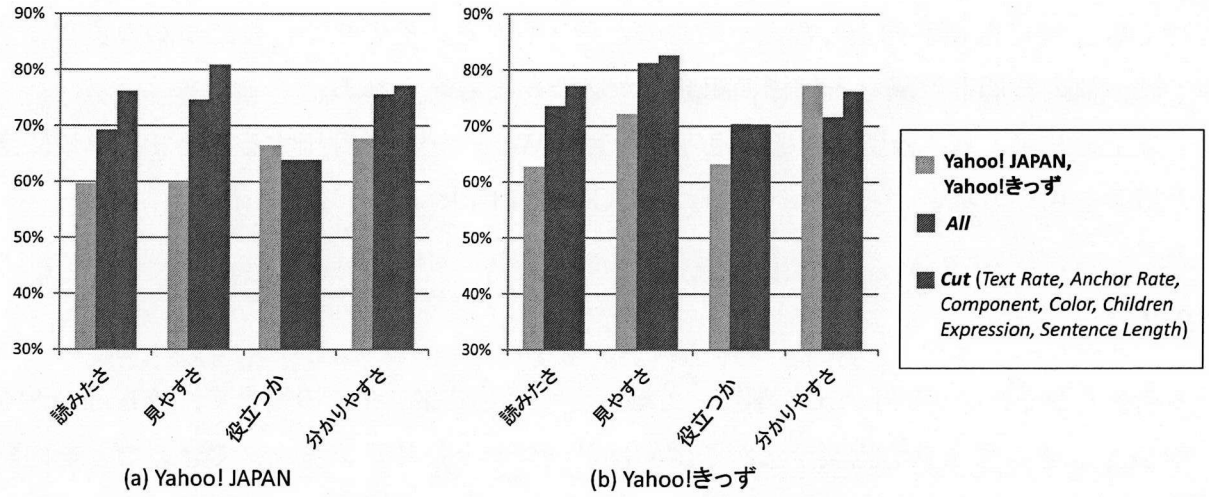


図 3.4: 指標を組み合わせた NDCG の平均値

この結果，*Cut* は，*Size*, *ImageRate*, *KanjiRate*, *Easy*, *DifficultExpression* の 5 つの指標を除いた，*TextRate*, *AnchorRate*, *Component*, *Color*, *ChildrenExpression*, *Sentence Length* の 6 つの指標の組合せとなる．

まず，全ての指標を組み合わせた *All* では，図 3.4(a) に示す Yahoo! JAPAN では，勉強になるかについては Yahoo! JAPAN と同程度，それ以外の質問で 5%から 15%の NDCG 値の向上が見られた．この結果より，単純に全ての指標を用いても，指標を組み合わせる効果があることが分かる．また，図 3.4(b) に示す，NDCG 値がもともと高い Yahoo!きっずでは，分かりやすさでは Yahoo!きっずの NDCG 値を下回ったが，それ以外の質問で 5%から 10%の NDCG 値の向上が見られた．分かりやすさで *All* の NDCG 値が下がってしまったのは，指標を単独で用いた場合でも Yahoo!きっずの NDCG 値を上回る指標がないため，それらを組み合わせることで性能が悪化したものと考えられる．

不要と考えられる指標をカットした *Cut* では，図 3.4(a) に示す Yahoo! JAPAN では，*All* の NDCG 値と同程度から 7%の向上が見られた．また，図 3.4(b) に示す Yahoo!きっずでも，同程度から 5%の NDCG 値の向上が見られた．この結果より，他の指標と相関しており，独立性の低い指標を除く効果が大きいことが分かる．ただし，役立つかの質問については，指標を除いたことによる NDCG 値の向上は見られなかった．これは，他の質問と異なり，単独で NDCG 値の向上する指標が *Color* や *ChildrenExpression* のみであり，他の

指標ではNDCG値が大きく下落するため，それらの指標の影響が大きく，指標を組み合わせても，NDCG値が向上しなかったものと考えられる．そのため，役立つかの質問については，他の質問とは異なる指標の組合せが必要だと考えられる．

以上の結果より，適切な指標を組み合わせてWebページを評価することで，単純に全ての指標を用いる場合よりに比べて概ねNDCG値を向上できることを確認した．

3.3.4 考察

指標をRanking SVMにより学習して組み合わせた結果，多くの場合で，Yahoo! JAPANやYahoo!きっずよりも提案手法が子供向けページを上位にランキングできることを確認した．しかし，Yahoo! JAPANの勉強になるかどうかでは，指標を組み合わせたNDCG値が0.65であるのに対して，*ChildrenExpression*を単独に用いたNDCG値が0.72であったように，単独指標の結果と比較して，指標を組み合わせることで性能が低下する場合があった．そのため，本項では，さらに精度を向上するために必要と考えられるクエリごとの特徴，また子供の学年ごとの特徴について考察し，本手法のさらなる改善方法について検討する．

クエリごとの特徴

Yahoo!きっずの分かりやすさのNDCG値が向上しなかった一つの原因はクエリによる差があることがあげられる．そこで，クエリごとにどのような特徴があるのかを考察する．

指標を単独を用いたリランク結果の中で，クエリによる差が顕著であった，Yahoo!きっずのデータセットにおける4つの指標 (*ImageRate*, *TextRate*, *ChildrenExpression*, *KanjiRate*) のクエリごとのNDCG値を，図3.5に示す．この結果，クエリごとに効果のある指標，効果のない指標に差が見られる．例えば，図3.5(c)のYahoo!きっずの分かりやすさで大きな差がみられ，“地球温暖化”や“冬至”はどの指標でもYahoo!きっずのNDCG値を下回っているが，“介助犬”や“百人一首”はどの指標も向上している．そのため，提案手法において，さらに精度を向上させるためには，クエリによる傾向の違いを考慮することが一つの方法として有効であると考えられる．具体的には，以下のような点を考慮する必要がある．

“ゲーム”，“うらない”という趣味に関するクエリは，図3.5(a)に示す読みたさや(b)に示す見やすさについて，他のクエリと比べて，NDCG値が下がったり，向上率が小さいと

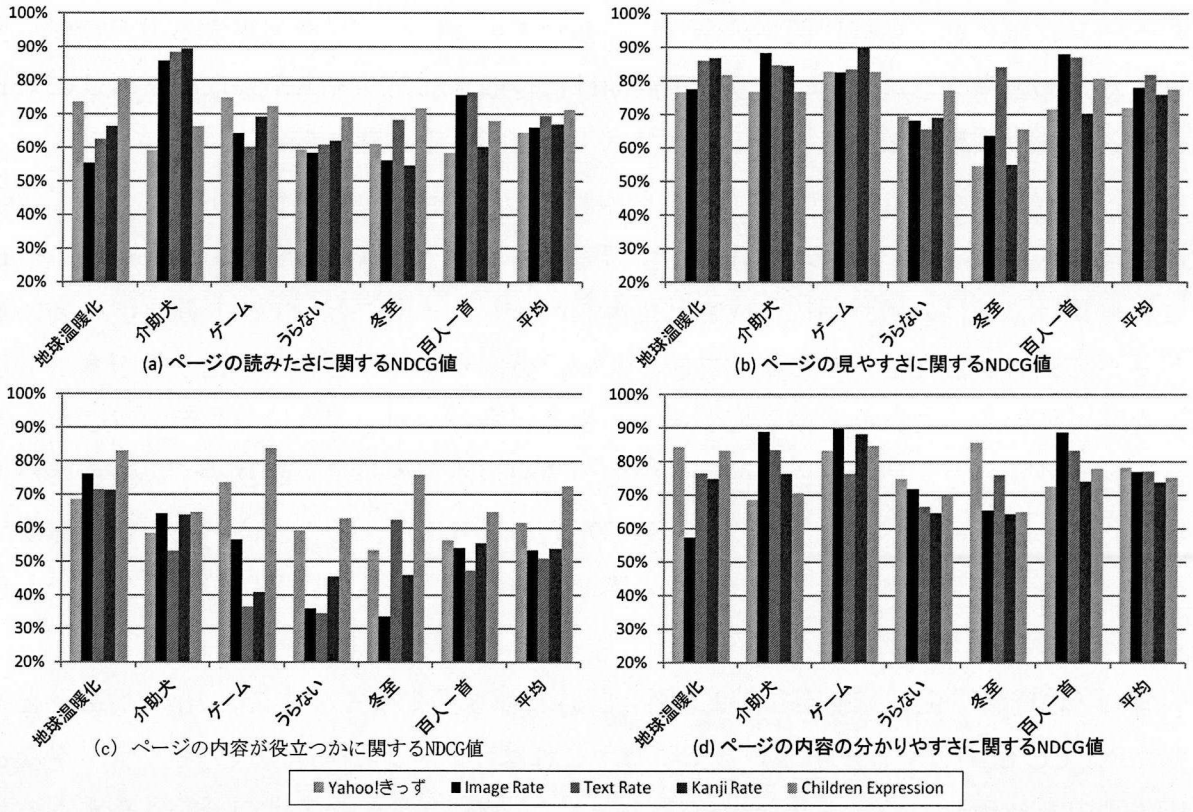


図 3.5: 各指標のクエリ別の NDCG の平均値 (Yahoo!きっず)

いう特徴が見られる．これは，個人の嗜好の影響が強いため，他のクエリと同じような組み合わせ方では NDCG 値が向上しないことが多いものと考えられる．例えば，今回の実験でページの評価をしてもらった被験者には男子が多かったため，サンリオやディズニーのページは全体的にスコアが低くなっていた．そのため，趣味に関するクエリについては，個人の嗜好に合わせた指標の組み合わせ方が必要になると考えられる．

“介助犬” のように子供にとって親しみのないクエリでは，図 3.5(a) に示す読みたさ，(b) に示す見やすさ，(d) に示す分かりやすさ全てにおいて，*ImageRate* や *TextRate* を考慮した NDCG 値が良く，他のクエリよりも NDCG 値の向上率が大きい傾向がみられる．このことより，子供にとって親しみのないクエリに対しては，ページに画像が多く含まれ，テキストが少ないことが子供の興味をひくために重要であるといえる．

また，“介助犬” のクエリの特徴の一つとして，他のクエリと異なり，*ChildrenExpression* よりも *KanjiRate* を考慮した方が NDCG 値が高い．これは，子供にとって親しみのない

クエリでは，子供向けに作成されたページが少なく，そのような中では子供向け表現を含むページは限られ，単純に漢字が余り用いられていないページを判定する方が精度が良くなったためと考えられる．つまり，子供向けに作成されたサイトが少ないクエリに対しては，子供向け表現の数を考慮するだけでは不十分であると考えられる．

図3.5(d)に示すYahoo!きっずの“地球温暖化”，“冬至”の分かりやすさについては全ての指標でNDCG値が下がり，(a)に示す読みたさでもNDCG値が下がる傾向がみられる．この結果は，子供の勉強用に作成された解説ページ，子供にとって少し難易度の高い解説ページの両方が混在し，その両方を読みたい，分かりやすいと見なした子供が多かったことを示している．これは，難易度が高めの文章が多いページでも，コンテンツごとに色分けをしたりすることで，どこに何についての情報があるのかが分かりやすい構成になっているため，子供は内容を分かりやすく感じたものと考える．そのため，ある程度難易度の高いページでも整理された分かりやすい構成であれば，子供は内容が分かりやすいと感じることを考慮する必要がある．

また“地球温暖化”，“冬至”では，図3.5(d)に示す分かりやすさで，*ImageRate*を考慮したNDCG値が20%程度も下がっている．これは，クエリと関連のないキャラクタなどの画像のみを使用しているようなページについて，特に高学年の子供の分かりやすさの評価が低かったことが原因と考えられる．この結果より，単純に画像を多く用いているだけでは，分かりやすさは不十分であり，クエリと関連している画像を用いることが子供にとっての分かりやすさに重要であると考えられる．そのため，子供向け解説ページでも，画像が直接的にクエリに関連しているものであるのかを考慮する必要がある．

以上の議論から，個人の嗜好の影響の大きい趣味に関するクエリ，検索エンジンのランキングに子供向け解説ページなどを多く含む学習に関するクエリ，子供向けページの少ない子供に親しみのない内容のクエリなど，クエリ特性に応じて指標の選択や組合せを決定することが，精度の向上に重要と考えられる．

学年ごとの特徴

データセットを作成するために，本評価実験に参加した小学生は60%が高学年を占めているが，学年により嗜好に差があるものと考えられる．そこで，この差を調べるために，図3.6にYahoo! JAPANをデータセットとした場合の単独指標のNDCG値を，小学校中学年，高学年それぞれで示す．Yahoo! JAPAN，Yahoo!きっずをデータセットとした両方

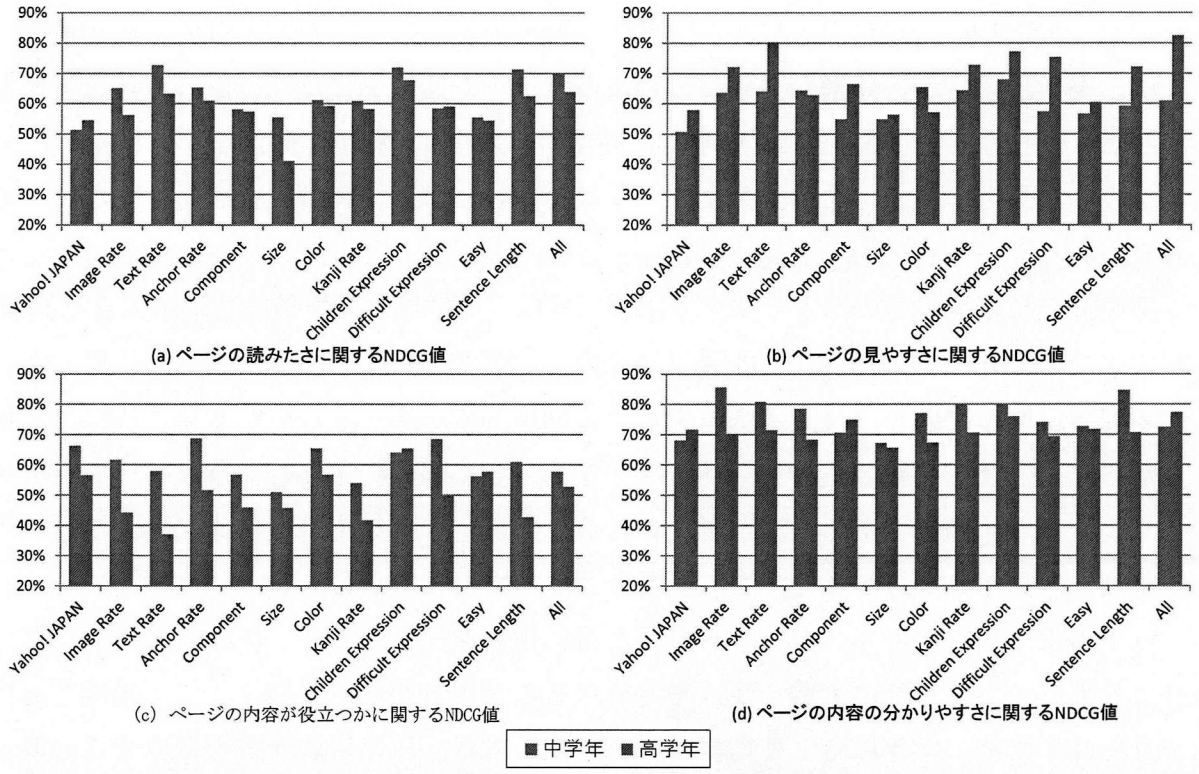


図 3.6: 学年別の NDCG 値 (Yahoo! JAPAN)

で，同様の特徴が見られたため，Yahoo!きっずをデータセットとした結果については省略する．また，低学年については，十分な数のページの評価を得られなかったため，結果を省略する．

図 3.6(a) に示す読みたさについては，中学年，高学年の結果は類似している．しかし，中学年の方が *Size*，*SentenceLength* の NDCG 値が高学年よりも高くなっている．これは，学年が上がるほど，Web ページを閲覧することに慣れているため，ページのサイズがある程度大きかったり，文章がある程度長かったりしても，読みたさにそこまで影響を与えなかったものと考えられる．それに対して，中学年ではページのサイズが小さく，一文も短い方がそのページを読みたいと思う傾向がある．

図 3.6(b) に示す見やすさについては，高学年では，構成に関する指標，文章に関する指標の両方の指標で NDCG 値の向上が見られるが，中学年では，構造に関する指標のみで NDCG 値の向上が見られる．この結果より，学年が上がるにつれて，ページの見た目だけでなく，書かれている文章にも影響を受けることが分かる．

図3.6(c)に示す役立つかどうかについては，高学年，低学年とも多くの指標でYahoo! JAPANの精度を下回っているが，中学年の方が構成に関する指標のNDCG値が高い．これは，高学年の方が，難解な文章が多いようなページを勉強になり役立つと見なす傾向があったためと考えられる．また，中学年は高学年と異なり，$DifficultExpression$のNDCG値が高い結果となった．この結果より，学年が小さいほど，難解な表現を好まないものと考えられる．

図3.6(d)に示す内容の分かりやすさについては，中学年と高学年で結果が大きく異なる．高学年ではほとんどの指標でYahoo! JAPANのNDCG値を下回っているのに対して，中学年では上回っており，提案手法が中学年に効果的であることが分かる．これは，学年が上がるほど，簡潔で親しみやすいページでなく，多くの説明を含むようなページを分かりやすいと見なす傾向があるためと考えられる．特に，中学年では，構成に関する指標の効果が高く，見た目が分かりやすさに影響を受けやすいことが分かる．

以上のように，中学年ほどページの見た目に影響を受けやすく，高学年になるほど見た目よりもページの中身に影響を受けやすくなることが分かる．このように，学年ごとに差が見られるため，学年に応じて指標のスコア化の方法，組み合わせ方を変更する必要があると考えられる．特に，分かりやすさについては，その影響が顕著であり，学年が上がるごとにページへの文章の理解度が増すことを考慮する必要がある．

3.4 むすび

本章では，子供のWeb検索を支援するために，検索結果を子供向けにリランクする手法を提案した．提案手法では，Webページの文章量や色の数などの構成に関する指標，文の長さや子供向け表現の数などの文章に関する指標をスコア化する．そして，各指標のスコアを組み合わせて各ページのスコアを決定し，スコアの降順に検索結果を並べ替える．

提案手法の有効性を検証するため，小学生の子供33人に，6個のクエリに関する300ページの評価を行ってもらったデータセットを作成し，そのデータセットを用いて評価実験を行った．その結果，子供向け表現を中心とした指標を組み合わせることで，子供にとっての“読みたさ”，“見た目の見やすさ”，“勉強の役に立つか”，“内容の分かりやすさ”の4つの側面において，提案手法によるリランクは，Yahoo! JAPAN，Yahoo!きっずのランキングよりNDCG値が最大20%向上した．この結果から，提案手法により，子供向け度合いの高いページを上位にランクでき，子供が検索結果から興味をひきやすいページを容易に選

択できるようになっていることが期待される.

第4章

Web検索クエリに関する子供向けサブトピックランキング手法

4.1 はじめに

第2章，第3章では，Webページ自身や検索結果のランキングを子供向けにすることにより，子供のWeb利用の支援を行うことを目的としている．しかし，子供はWeb検索の際にクエリの入力を苦手とするという報告があるため，検索結果やWebページだけでなく，検索方法の支援も重要である [16].

Web検索方法の支援の有効な方法として，クエリに関するサブトピックを利用する研究が盛んに行われている．クエリに関するサブトピックとは，入力されたクエリに対して“ユーザの意図を具体化する文字列”である．例えば，“ハリー・ポッター”というクエリについては，“映画”，“キャラクタ”などがユーザの意図を示すサブトピックの候補として考えられる．これらのサブトピックを利用したクエリサジェスション，検索結果のクラスタリング，ランキングの多様化などにより，ユーザのWeb検索を支援可能となる．

子供にとっても，サブトピックを利用した支援は有効であると考えられる．しかし，既存研究のサブトピック抽出方法 [14,35,37,38,76] において，検索エンジンの検索結果やクエリログ，外部ソースから取得できる一般的に重要なサブトピックが，子供にとっても重要であるとは限らない．

文献 [2] で，子供にとっては，興味を持って楽しむことが学習に繋がり，重要であるという指摘があるように，子供のWeb検索のためにも，子供にとって興味をひくサブトピック

を提示することが重要であると考えられる．さらに，子供にとって重要なサブトピックは，子供の検索目的を考慮すると，“趣味”や“学習”に役立つものであると考えられる[1]．例えば，クエリ“沖縄”については，“ツアー”，“レンタカー”というような旅行に関するサブトピックが一般的に重要と考えられるが，子供自ら沖縄旅行をする際に，ツアーやレンタカーは子供にとって興味をひく内容ではなく，それらに関する情報を調べる状況は考えにくい．一方，“食べ物”，“歴史”というようなサブトピックは，子供が授業で沖縄の文化や習慣について学んでいる際に興味をひく内容であり，重要であると考えられる．子供の興味をひくサブトピックを提示することで，そのサブトピックについて調べてみたいと子供の欲求を刺激し，Webを活用できるきっかけになると考えられる．もちろん，子供の興味をひくサブトピックが必ずしもWeb検索の際に役に立つとは限らないが，まず興味をひくものでなければ，中身を調べようという意欲が沸きにくい．つまり，Web検索の際に子供にとって重要なサブトピックは，子供にとって興味をひき，役に立ちそうと感じるものである上で，Web検索に実際に役に立つものである必要がある．そのため，子供のWeb検索支援のために重要なサブトピックを特定するには，何らかの方法で，子供目線での重要度とWeb検索時に役立つ重要度の両方の面を定量化しなければならないと考えられる．

そこで本章では，まず，子供目線で興味をひくかどうかという観点でのサブトピックの重要度を定量化することに着目し，子供にとって重要なサブトピックを優先的に提供するため，1) 子供向けサブトピックの調査，および，2) 子供向けサブトピックランキング手法の検討について取り組む．子供の興味をひくという点では，まず，全く内容を想像できないような分かりにくいサブトピックについては，検索クエリとの関係性が分からず，そのサブトピックについて興味を持ちにくいと考えられるため，サブトピック自身が子供にとって分かりやすいかどうかが一つの重要な要素となり得る．また，興味をひくサブトピックは，実際にWebで調べてみたいと感じると考えられるため，サブトピック自身を調べてみたいかどうかも重要な要素となり得る．さらに，検索クエリについて調べている際に，そのサブトピックの内容について知ることが役に立つと感じるのかどうかも，サブトピックへの興味を左右する重要な要素と考えられる．そのため本章では，子供にとって分かりやすい，調べてみたい，役に立ちそうと感じるサブトピックを子供向けサブトピックとする．

具体的には，子供向けサブトピックの調査に関しては，一般向け検索エンジン，子供向け検索エンジン，Wikipediaという複数のソースから取得したサブトピックを，分かりや

[1]<NTT レゾナント：http://research.goo.ne.jp/>

すさ，調べたさ，役立つかどうかの3つの観点からスコア付けしてもらうことで行う．そして，高スコアのサブトピックを取得できるクエリやソースの特徴，難易度などのサブトピックの要素とスコアの関係，学年ごとのスコアの比較という観点からスコア付けしてもらった結果の調査を行う．調査の結果，以下のような点が子供向けサブトピックの重要度の定量化に重要であることが明らかになった．

- 子供向けサブトピックを簡単に取得できるのは検索エンジンの検索結果から得られる関連検索語であるが，重要なサブトピックを広くカバーするには，Wikipediaなどの他のソースも効果的である．

- サブトピックの抽象度，難易度，検索意図，検索ヒット数といった要素を考慮することが，子供向けサブトピックを特定するための手がかりとなる．具体的には，学習を検索目的とする場合には，沖縄の“歴史”といったクエリに関する一般名詞によるサブテーマ，趣味を検索目的とする場合には，“動画”といった情報の形式に関する検索意図を示し，抽象度がある程度高く，難易度が高くないサブトピックを優先的に提供するのが効果的である．

- 学年ごとに重要視するサブトピックの特徴を変化させるべきである．学年が下がるほど，サブトピックの難易度や抽象度への重要度が増す．

以上のような知見に基づいて，さらに本章では，子供向けサブトピックを優先的に提供するための子供向けサブトピックランキング手法について検討する．子供向けサブトピックランキング手法では，知見に基づいて決定した取得ソースの種類や難易度などのサブトピックの要素に関する指標をもとに，サブトピックの子供向け度合いを数値化することで行う．これにより，子供が検索クエリを入力した際に，子供向けサブトピックをクエリサジェスションとして提示したり，子供向けサブトピックに基づいて検索結果のクラスタリングを行うことで，子供は自分の検索要求を細かいクエリとして入力する手間なしに，必要とするサブトピックについての情報を容易に得ることできる．

以下では，まず4.2節でサブトピック抽出に関する関連研究について述べ，4.3節で子供向けサブトピックの調査について述べる．その後，4.4節で子供向けサブトピックランキング手法について述べ，最後に4.5節で本章のまとめを行う．

4.2 関連研究

子供向けではないが，サブトピックを抽出し，利用する研究が近年多数行われている．文献 [35] では，クリックスルーログ，Web 検索結果で返されたドキュメント内のアンカーテキスト，n-グラムを用いて，サブトピックを抽出し，検索結果の多様化を行っている．文献 [38] では，クエリとクリックスルーのログを用いて，クエリとクリックされたドキュメントからなるグラフをクラスタリングし，サブトピックを抽出している．文献 [76] では，クエリの属性に応じたテンプレートをクエリログから作成して利用することで，サブトピックを抽出している．文献 [14] では，クエリログ，アンカーテキスト，検索結果中のキーフレーズを元に，サブトピックを抽出し，検索結果の多様化を行っている．文献 [37] では，クエリを Wikipedia の概念にマッピングして利用することで，クエリの意図を推定し，カテゴライズを行っている．

以上のように，これらの研究では，Web 検索エンジンから返される検索結果，外部ソース，クエリやクリックスルーのログなどを利用して，サブトピックを抽出する．しかし，これらの手法で取得できるサブトピックは子供向けとは限らないため，子供にとって重要なサブトピックについて調査する必要がある．本研究では，これらの既存研究で様々なソースからサブトピックを抽出している点に着目し，様々なソースから抽出したサブトピックを子供に評価してもらうことで，子供の興味をひき，役に立つサブトピックを抽出するのに各ソースをどのように扱うべきかについて検討する．

4.3 子供向けサブトピックの調査

4.3.1 サブトピックのスコア付け

データセット

子供が検索する可能性の高いクエリで調査を行うため，Yahoo!きっずの 2009 年 12 月から 2011 年 10 月の週間検索キーワードランキング上位 10 位より選んだ “地球温暖化”，“聖徳太子”，“金環日食”，“沖縄”，“恐竜”，“トマト”，“ポケモン”，“ハリー・ポッター” の 8 個のクエリを想定した．

クエリには，学習を主な目的とするもの，趣味を主な目的とするもの，学習と趣味のどちらも目的とし得るものが存在し，何を目的とするかによって重要なサブトピックが異な

ると考えられる．そのため，“地球温暖化”，“聖徳太子”，“金環日食” については学習のみ（3ペア），“沖縄”，“恐竜”，“トマト” については学習と趣味の2種類（6ペア），“ポケモン”，“ハリー・ポッター” については趣味のみ（2ペア）の検索目的を想定した合計11種類の検索目的とクエリのペアについてサブトピックを抽出し，データセットとした．

サブトピックを抽出するには，検索エンジンの検索結果，Wikipediaなどの外部ソース，クエリやクリックスルーのログを利用する方法が代表的である．そこで，サブトピックの抽出には，子供がWeb検索の際に利用する一般向け検索エンジンや子供向け検索エンジンにおける関連検索語や検索結果，代表的な外部リソースとしてWikipediaを利用した．具体的には，以下の6種類のソースからサブトピックを抽出した．

- genRel：Yahoo! JAPANにおけるクエリの関連検索語
- genKey：Yahoo! JAPANにおけるクエリの検索結果上位100件のスニペットに出現するキーフレーズ
- kidRel：Yahoo!きっずにおけるクエリの関連検索語
- kidKey：Yahoo!きっずにおけるクエリの検索結果上位100件のスニペットに出現するキーフレーズ
- kidCat：Yahoo!きっずにおけるクエリの検索結果上位100件の登録カテゴリ
- wiki：クエリをタイトルとするWikipediaの記事内の項目タイトル

キーフレーズについてはキーフレーズ抽出API[2]で取得した．このAPIを用いることで，検索結果のスニペットの文章を解析して特徴的な表現とその重要度を取得できる．

関連検索語からは他のソースと比較すると抽出できるサブトピックは少数であるが，ユーザが頻繁にクエリと一緒に検索する語であるため，サブトピックとしての重要性が高いと考えられる．例えば，検索エンジンの検索結果に提示される関連検索語は，Yahoo! JAPANで最大10個，Yahoo!きっずで最大3個である．一方，検索結果中のキーフレーズについては，非常に多数のサブトピックを抽出することが可能だが，ノイズも多く含まれると考えられる．Yahoo! JAPAN，Yahoo!きっずの両方を利用したのは，Yahoo! JAPANとYahoo!

[2]<Yahoo!デベロッパー キーフレーズ抽出API: http://developer.yahoo.co.jp/webapi/jlp/keyphrase/v1/extract.html>

きっずでは検索結果に含まれるページが異なるためである．Yahoo!きっずの検索結果では，クエリに対応するお薦めサイトがあれば，そのサイトを検索結果上位に表示し，その後，Yahoo! JAPANと同様のページ検索結果のランキングを表示するが，あらかじめ登録されていないページは表示されないため，Wikipediaなどの子供向けでないと考えられるページが少ない．また，Yahoo!きっずのおすすめサイトにはそれぞれカテゴリが登録されており，検索結果中に表示される．カテゴリはクエリの大まかな意図を示すと考えられるため，サブトピックを取得する一つのソースとして用いた．カテゴリは具体的ではないため，意図を大幅に絞ることはできないが，子供にとって知っている語が多いと考えられる．Wikipediaについては，記事全体のキーフレーズは子供にとって難解である可能性が高いが，項目タイトルはそこまで難解でなく，記事タイトルであるクエリの内容を説明するサブトピックになり得ると考えられる．

複数ソースから取得したサブトピック中からクエリに該当する文字列は除外して統合し（例："沖縄旅行" と "旅行" は "旅行" に統合），複数のソースから同一のサブトピックを取得した場合も1つのサブトピックとして扱った．そして，Yahoo!キーフレーズAPIで返ってきた重要度のスコアが低いサブトピックは除外し，各クエリのサブトピックを100個とした．

手順

評価実験のために，Web上で簡潔な操作で評価を行えるシステム（図4.1）を構築した．被験者は，小1から中3の複数の学校に所属する子供49名とし，データセット内の3つの質問によるサブトピックの評価を行ってもらった．被験者の内訳は小学校低学年が1名（男1名），小学校中学年が11名（男5名，女6名），小学校高学年が25名（男19名，女6名），中学生が12名（男7名，女5名）であった．3つの質問で評価を行ったのは，内容を少しでも知っているのかどうか，検索したい意欲を促進するのかどうか，検索時に実際に役立ちそうなのか，という多面的な要素で，子供の興味をひき，重要なサブトピックがどのようなものなのか評価を行うためである．以下に実験の手順を示す．

1. 指定されたクエリ q と検索状況 *situation* で連想できるサブトピックを自由に入力してもらう．

2. ランダムに表示されたサブトピック s について，以下の質問に'はい'，'どちらでもない'，'いいえ'で回答してもらう．

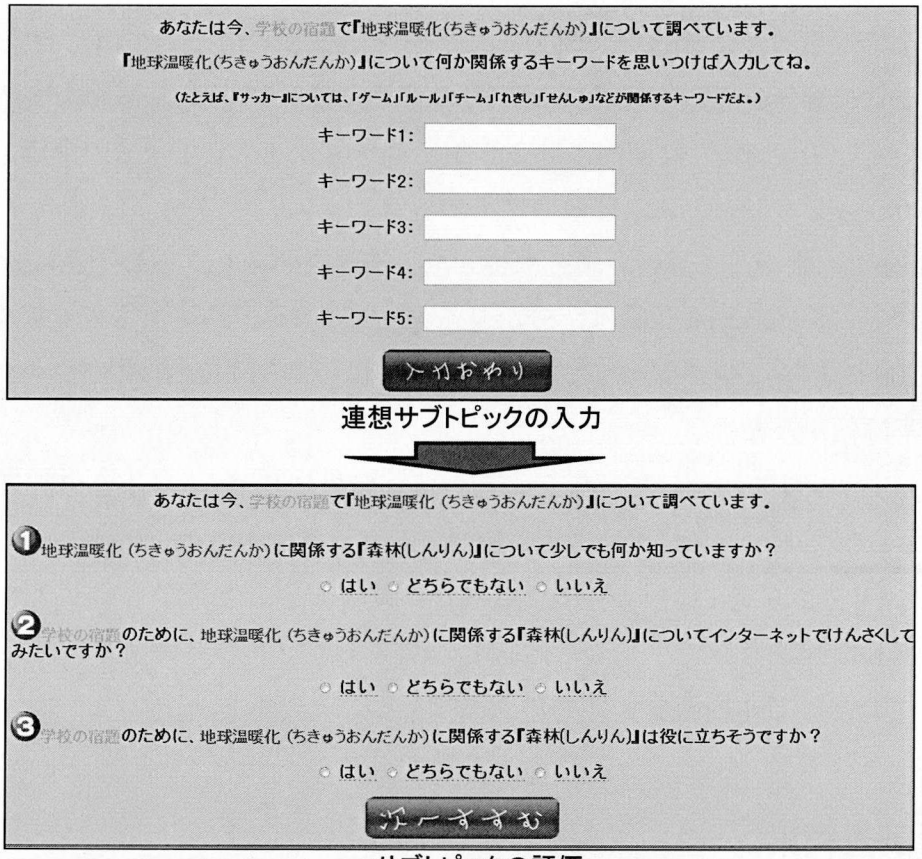


図 4.1: サブトピック評価システム

- q に関係する s について少しでも何か知っていますか？(分かりやすさ)
- *situation* のために，q に関係する s についてインターネットで検索してみたいですか？(調べたさ)
- *situation* のために，q に関係する s は役に立ちそうですか？(役立つかどうか)

3. 各クエリで (2) を複数サブトピックで繰り返す.

4. (1) から (3) を複数クエリで繰り返す.

検索状況 *situation* は，学習を目的とする “学校の宿題”，趣味を目的とする “友達と遊ぶ” などのあらかじめ定義した状況を利用する．“地球温暖化”，“聖徳太子”，“金環日食” については学習，“沖縄”，“恐竜”，“トマト” については学習と趣味，“ポケモン”，“ハリー・ポッター” については趣味を目的とする検索状況を用意した.

評価については，最初に，被験者に大学に集まってもらい，1日で100個のサブトピックを評価してもらった．その後，協力してもらえる被験者については，自宅からも自由にサブトピックの評価を行ってもらった．その結果，各サブトピックに対して，少なくとも4人分の評価を取得した．具体的には245個のサブトピックで4人分の評価，455個のサブトピックで5人分の評価，320個のサブトピックで6人分の評価，77個のサブトピックで7人分の評価の合計5,714個の評価を取得した．被験者間のサブトピックに対する評価の一致率を示すFleiss's Kappa [26]は，分かりやすさで0.234，調べたさで0.208，役立つかどうかで0.206であり，サブトピックに対する評価は年齢による差が大きいと考えられるため，一致率は低かった．

その後，3つの質問について評価してもらった結果に基づき，1,100の評価用サブトピックに対して，サブトピックごとのスコアを決定した．具体的には，評価の'はい'を1点，'どちらでもない'を0.5点，'いいえ'を0点とし，質問ごとに各サブトピックの合計点数を算出し，評価した被験者数で平均した値をそのサブトピックのスコアとした．スコアが高いほど，子供にとって高評価なサブトピックであることを示す．

4.3.2 調査結果

4.3.1項で述べたサブトピックのスコアの傾向を調査することで，子供の興味をひくサブトピックを優先的に提供するために必要な要素について検討する．まず，分かりやすさ，調べたさ，役立つかという3つの質問間の関係について調査し，子供にとって興味を持つサブトピックの全体像を明らかにする（調査項目1）．次に，子供が評価したサブトピックのスコアに影響する要素について調査し，子供向けサブトピックを特定するために重要な点を検討する．具体的には，まず，データセットとして用意したクエリやサブトピックの抽出元ソースの種類が，抽出できるサブトピックの種類に影響し，子供によるサブトピックのスコアに影響すると考えられるため，クエリやソースの種類とサブトピックのスコアの関係について調査する（調査項目2）．これにより，検索目的の違いによるサブトピックのスコアの違いや子供向けサブトピックの抽出に重要なソースを明らかにする．また，クエリやソースの種類を考慮しただけでは，子供向けサブトピックとそれ以外を区別するためのサブトピックの定量化が難しいと考えられるため，子供が重要視すると予想されるサブトピック自身の抽象度や検索意図などの特徴とサブトピックのスコアの関係について調査する（調査項目3）．さらに，子供は，学年により，学校で学ぶ内容や趣味が大きく変化す

ると考えられるため，学年による違いを考慮することが，子供向けサブトピックの特定に重要な要素の一つと考えられる．そのため，学年ごとのサブトピックのスコアの差，重視するサブトピックの特徴の差についても調査し，学年の差を考慮するべきかどうかを明らかにする（調査項目 4）．最後に，子供にサブトピックをスコア付けしてもらう際に自由に入力してもらったサブトピックの内容について調査し，データセットとして用意した抽出サブトピックとの差について調べ，子供向けサブトピック提示の重要性を明らかにする（調査項目 5）．

このような調査の結果，子供向けサブトピックの抽出に重要であると考えられる以下の点が明らかになった．

1. 学習を目的とするか趣味を目的とするかで重要なサブトピックが変化する．
   趣味に関するクエリでは嗜好の影響が強く出る（調査項目 2）．一方，学習に関するクエリは，子供にとって一見親しみにくくとも，役に立ちそうな重要なサブトピックが多数あり（調査項目 1，調査項目 2）．また，検索目的の違いによって，必要とされる意図は異なる（調査項目 3）．学習に関するクエリでは，クエリに関するサブテーマを示すような一般名詞が高スコアであり，趣味に関するクエリでは動画像のような情報形式に関するサブトピックが高スコアである．

2. 幅広く重要な意図をカバーするためには，関連検索語を中心に複数のソースを利用すべきである．
   検索エンジンの関連検索語から抽出するサブトピックが高スコアであるが，数が少なく，カバーできる意図が少ない（調査項目 2，調査項目 3）．検索結果中のキーフレーズや Wikipedia というソースを利用すると，難易度が高くなってしまうが，関連検索語だけでは抽出できない調べたい，役に立つと感じるサブトピックを数多く抽出することができると考えられる（調査項目 2）．さらに，一般向け検索エンジンと子供向け検索エンジンというソースの種類で抽出できるサブトピックの示す意図の種類が異なるため，重要な意図を幅広くカバーするためには，複数ソースを利用すべきと考えられる（調査項目 3）．

3. サブトピックの抽象度，難易度，意図，検索ヒット数といった要素は子供向けサブトピックを特定するための手がかりとなる（調査項目 3）．
   複数の要素を組み合わせることで，子供向けサブトピックの定量化を効果的に行える

と考えられる．ただし，抽象度の低い固有名詞のサブトピックは子供にとって知っているものであるかどうかが重要となる．

4. 学年ごとに重要視するサブトピックの要素は一定ではない（調査項目 4）．
学年が上がるほどサブトピックへのスコアは高くなるが，検索目的やソースの違いは学年に関わらず重要である．また，全ての学年でサブトピックの示す意図や検索ヒット数を考慮すると効果的だと考えられるが，難易度や抽象度は中学生になるとそこまで重要でなくなる．

5. 子供自身の力で連想できるサブトピックには限界がある（調査項目 5）．

以下では，これらの知見の詳細について順に述べる．なお，上記において，括弧内が各知見に関連する結果について述べる調査項目に対応している．

1. スコア同士の関係

まず，3 つの質問に対する評価が独立しているのか，相関しているのかを調査するために，同じサブトピックに対する各質問のスコア間のピアソンの相関係数を算出した．その結果，分かりやすさと調べたさの間は 0.345，分かりやすさと役立つかどうかの間は 0.391，調べたさと役立つかどうかの間は 0.732 となった．この結果より，全ての質問間でサブトピックに対する評価はやや相関しているが，特に調べたさと役立つかどうかは高い相関にあることが分かる．これは，サブトピックを知らないから調べたくないという評価よりも，あまり知らないから，学習になって役に立ちそうで，積極的に調べてみたいという評価が多かったためであると考えられる．つまり，一見親しみにくくとも，子供にとって調べたい，役に立つと感じる重要なサブトピックは多数あることが分かる．

2. クエリやソースとスコアの関係

クエリとソースによるサブトピックのスコアの違いについての調査結果を述べる．クエリの各検索目的でのスコアの平均値を表 4.1 に，各ソースでのスコアの平均値を表 4.2 に示す．また，各ソースで抽出したサブトピックの個数を表 4.3 に示す．サブトピックのスコアに対して，0.05 を有意水準とし，クエリとソースの 2 要因の分散分析を行ったところ，(1) 分かりやすさについてはクエリ，ソース，(2) 調べたさについてはクエリ，(3) 役立つかどうかについてはクエリ，ソースの違いにより，有意にスコアが変化していることが分かった．

表 4.1: クエリの各検索目的の平均スコア

	学習	趣味
分かりやすさ	0.376	0.401
調べたさ	0.538	0.493
役立つかどうか	0.541	0.586

表 4.2: 各ソースの平均スコア

	wiki	genRel	genKey	kidRel	kidKey	kidCat
分かりやすさ	0.371	0.506	0.407	0.684	0.397	0.427
調べたさ	0.531	0.562	0.508	0.615	0.519	0.515
役立つかどうか	0.537	0.565	0.513	0.668	0.518	0.515

この結果より，クエリやソースの違いが子供にとってのサブトピックの評価と相関していることが分かる．さらに，分散分析 [3] の後，有意な要因について Bonferroni 法 [20–22] で多重比較を行い，0.05 の有意水準で有意に平均値が異なる組合せを調査した．

クエリごとの違いについては，多重比較の結果，“ポケモン” や “ハリー・ポッター” という趣味に関するクエリが，“地球温暖化” や “聖徳太子” という学習に関するクエリとスコアの平均値が有意に異なることが分かった．具体的には，分かりやすさについては，学校で学んだ内容かどうかに依存する学習に関するクエリのスコアが，趣味に関するクエリのスコアよりも低かった．調べたさや役立つかどうかについては，趣味，特に “ポケモン” や “ハリー・ポッター” という遊ぶという検索目的下では，嗜好の影響が強く出るため，興味のないクエリに関するサブトピックのスコアを低く評価する子供が多かった．そのため，趣味に関するクエリのスコアは，学習に関するクエリのスコアを下回った．この結果より，学習を目的とした際には，子供にとって，親しみが薄くとも，調べてみたい，役に立ちそうと感じるサブトピックが多く存在すると考えられる．

ソースごとの違いについては，多重比較の結果，分かりやすさについては genRel と kidRel，役立つかどうかについては kidRel のみが他のソースのスコアの平均値を有意に上回った．これは，検索エンジンの関連検索語は実際にユーザが頻繁に検索しているキーワードの組

表 4.3: 各ソースから抽出したサブトピックの個数

wiki	genRel	genKey	kidRel	kidKey	kidCat
274	106	378	28	507	190

合せであり，子供にとっても分かりやすいサブトピックが多く含まれるためと考えられる．genRel と kidRel を比較すると，Yahoo! JAPAN は一般向け検索エンジンであり，大多数のユーザは大人であるため，genRel の中には "沖縄ツーリスト" など子供にとって役に立たないと感じるサブトピックが多くあり，genRel のスコアは kidRel のスコアを下回った．このように関連検索語を取得できる際には，積極的に子供向けサブトピックとして利用することが効果的であると考えられる．ただし，関連検索語から取得できるサブトピックは非常に少ないため，幅広いサブトピックを取得するためには，他のソースも利用すべきと考えられる．

他のソースの傾向としては，wiki では，難解な表現ではあるが，学習に役立ちそうなサブトピックが多いと考えられるため，分かりやすさは全てのソースの中で最低であるが，調べたさや役立つかどうかは比較的高かった．つまり，調べたさや役立つかどうかを重視する際には，Wikipedia などのやや難解ではあるが整理された外部情報を利用することが効果的であると考えられる．genKey や kidKey からは，数多くのサブトピックを抽出できるが，サイト名や団体名などの子供にとって想像がつかないであろう固有名詞を多く含むと考えられるため，全体として平均スコアは高くなかった．そのため，多様なサブトピックを抽出するためには，検索エンジンを利用することが必要になるが，単純に重要語を抽出しただけでは不十分であり，子供にとって明らかに不要なサブトピックを排除する必要がある．kidCat は，固有名詞や難解な表現は少ない分，分かりやすさは wiki や kidKey よりもやや高かった一方，抽象度の高いカテゴリは曖昧であり，検索時に役立つイメージがつきにくいためか，調べたさや役立つかどうかというスコアは高くなかった．つまり，カテゴリについても，抽象度が高くて役に立たないなど，不要なサブトピックを排除する必要がある．

以上の結果より，子供にとって重要なサブトピックを幅広く抽出するには，子供向け関連検索語を中心に，その他のソースも利用してサブトピック候補を抽出した後，何らかの特徴を考慮して，子供向けのサブトピックとしての重要性を決定する必要があると考えら

れる.

3. サブトピックの要素とスコアの関係

子供向けサブトピックとしての重要性を定量化するために，サブトピックのスコアに影響を与えると考えられるサブトピックの要素について検討する.

子供は語彙が少ないため，子供向け文章には漢字やアルファベットの割合や文字数が少ない．そのため，漢字数やアルファベット数が子供によるサブトピックの評価に影響すると考えられる．また，サブトピックの抽象度もサブトピックの評価に影響すると考えられる．つまり，具体的なサブトピックには難解なものが多くなるので，評価が低くなる可能性がある.

さらに，サブトピックが示す検索意図が重要となる．一般的に，検索を行う際には，どのような情報要求を達成したいかという何らかの検索意図が存在し，その検索意図に対応し，その検索に有効なサブトピックが存在すると考えられる．つまり，提示するサブトピックは検索意図と対応すべきである．これをサブトピックの示す検索意図とする．例えば，“沖縄” というクエリについては，“レンタカー”，“ホテル” というサブトピックは旅行プランに関する検索意図，“戦争”，“年表” というサブトピックは歴史に関する検索意図を示すと考えられる．サブトピック自体が分かりやすくとも，サブトピックの示す意図と子供にとっての情報要求が食い違えば，重要なサブトピックとはならないため，サブトピックの示す意図がどのようなものであるのかも重要であると考えられる.

さらに，サブトピックを示すキーワードを用いて，検索エンジンで検索した際に返ってくる検索ヒット数も，子供にとって重要なサブトピックなのかどうかを示す一つの要素になると考えられる．検索ヒット数の極端に少ないようなサブトピックは，一般的に利用されにくい不要なサブトピックである可能性が考えられる．以上の議論に基づいて，子供のサブトピックの評価に影響する要素として，本研究では以下の 4 点について調査する.

- difficulty：サブトピックの難易度．漢字やアルファベットの文字数.
- abstraction：サブトピックの抽象度（1：低〜5：高）.
- intent：サブトピックの示す検索意図.
- genHit (kidHit): Yahoo! JAPAN（Yahoo!きっず）でサブトピックで検索して返ってきたページ数.

表 4.4: 各ソースから抽出したサブトピックの要素の平均値

	wiki	genRel	genKey	kidRel	kidKey	kidCat
intent 数	9.72	4.72	13.454	2.081	13.909	9.000
difficulty	2.916	1.971	2.592	1.821	2.441	2.126
abstraction	3.120	2.789	2.652	3.523	2.940	3.380

表 4.5: 各ソースで抽出したサブトピックが表現している意図の例

	地球温暖化	沖縄	ポケモン
wiki	影響, 対策, 議論	社会, 文化, 自然	ゲーム, テレビ, 市場
genRel	対策, 概要	ツアー, サイト	ゲーム, サイト
genKey	対策, 問題, 影響	ツアー, 旅行, 暮らし	ゲーム, キャラ, サイト
kidRel	問題, 原因	文化, 料理	ゲーム
kidKey	対策, 問題, 団体	社会, スポット	ゲーム, キャラ, 団体
kidCat	団体, 対策	社会, 文化	ゲーム, テレビ, 映画

difficulty については，本調査では，子供にとって難解な文字であると考えられる漢字，アルファベットの合計文字数とした．difficulty と intent については，3名の評価者で全サブトピックについてアノテーションを行った．具体的には，abstraction については3名の評価者が，5段階評価で各サブトピックの抽象度を付加し，3名の評価者に付加された抽象度の平均値をそのサブトピックの abstraction とした．intent については，3名の評価者が，各クエリのサブトピック 100 個を 10 個から 15 個の意図になるように分類し，3名の評価者に付加された意図のうち，大多数を占める意図をそのサブトピックの intent とした．intent は，“サイト名”，“概要情報” などの複数クエリで共通の意図，クエリ “沖縄” についての “旅行”，クエリ “金環日食” についての “観測” などの単独クエリのみに存在する意図の両方を合わせて，計 85 種類となった．3名の評価者の一致度を示す Fleiss's Kappa は，抽象度で 0.701，intent で 0.672 あり，中から高程度一致していた．

まず，取得元ソースの特徴を調べるため，各ソースのサブトピックの要素について平均値を算出した（表 4.4）．intent 数とは，各ソースで抽出したサブトピックに付与した意図

の数のクエリごとの平均値である．

difficultyについては，wikiやgenKey，kidKeyという分かりやすさのスコアが低いソースで値が高くなった．この結果より，Wikipediaや検索エンジンのサブトピックは漢字やアルファベットが多く，子供にとって親しみにくかったと考えられる．abstractionについては，よく検索されるサイト名を含むgenRel，サイト名，人名を多く含むgenKey，kidKeyで比較的低く，子供に理解しやすい表現であるkidRelやkidCatで比較的高いものが多かった．abstractionの低いgenKeyやkidKeyの分かりやすさのスコアはあまり高くなかった一方，同じようにabstractionの低いgenRelのスコアは高かった．これは，genRelのサブトピックはユーザがよく検索する語であるため，子供にとってもある程度身近で想像でき，評価が高かったものと考えられる．この結果より，具体的なサブトピックについては，子供にとって知っているものであるかどうかが重要であると考えられる．

intentについては，kidRelやgenRelではintentの数が他のソースに比べて大幅に少なく，関連検索語のみからではサブトピックの意図が狭まってしまうことが分かる．具体的に，各ソースで抽出したサブトピックがどのような意図を表現するのかを表4.5に示す．表4.5では，3種類のクエリについて，該当するサブトピック数の多い代表的な意図について示している．genRelやkidRelだけでは，一部の意図のみを示すサブトピックしか取得できないことが分かる．

また，クエリの種類の違いが各ソースから抽出できる意図の種類に影響していることが分かる．特徴的な意図としては，genRelやgenKeyでは“サイト”といった意図，kidKeyやkidCatでは“団体”といった意図を示す他では抽出できないサブトピックが多数存在した．また，学習に関するクエリ“地球温暖化”，趣味に関するクエリ“ポケモン”については，異なるソースでも同じような意図を示すサブトピックを抽出できる一方，学習と趣味に関するクエリ“沖縄”については，子供向け検索エンジンと一般向け検索エンジンで異なる意図に対応するサブトピックを抽出できることが分かる．具体的には，前者からは“旅行”や“ツアー”という旅行の情報収集，後者では“文化”や“社会”という沖縄の暮らしの学習という意図がサブトピックの大部分を占めていた．

さらに，より直接的に，各要素とサブトピックのスコアの関係を調べるために，各要素の値と分かりやすさ，調べたさ，役立つかどうかの相関係数を求めた（表4.6）．表4.6に示すように，difficultyは分かりやすさと弱い負の相関，intentは分かりやすさと弱い正の相関，genHitとkidHitは分かりやすさと役立つかどうかとの間に弱い正の相関，intentに

表 4.6: サブトピックの要素とスコアの相関係数

	intent	difficulty	abstraction	kidHit	genHit
分かりやすさ	0.447	−0.284	0.294	0.286	0.252
調べたさ	0.370	0.005	0.133	0.171	0.170
役立つかどうか	0.380	−0.019	0.197	0.231	0.202

表 4.7: 高スコアな意図とサブトピック例（意図の例(サブトピック例:取得元ソース)）

地球温暖化	沖縄(学習)	沖縄(趣味)	ポケモン
現状 (現状:genRel,genKey)	歴史 (米軍基地:wiki)	スポット(美ら海水族館: genKey,kidKey)	カード(カードゲーム: genKey,kidKey)
乗り物 (乗りもの:kidCat)	地名 (那覇市:genKey)	気候 (気候:wiki,kidRel)	動画像 (イラスト:genKey)
影響(気候変動: wiki,genKey)	クイズ (クイズ:KidKey)	動画像 (地図:genKey)	団体 (任天堂:kidKey)

ついては分かりやすさ，調べたさ，役立つかどうかと弱から中程度の正の相関が見られた．これらの相関は単独では十分に大きいものであるとはいえないが，これらの要素を組み合わせることで，子供向けサブトピックの定量化の際に有効な指標になり得ると考えられる．特に意図については考慮する必要があると考えられる．

そこで，具体的に，どのような意図が子供にとって高評価なのか，3種類のクエリの例を表4.7，4.8に示す．各クエリで，高（低）スコアな意図3つとそれぞれの意図の中で最高スコアのサブトピック，その抽出元のソースを括弧内に示す．

学習に関するクエリに対しては，“動画像”，“リンク集”という情報の形式に関する意図が低スコアである一方，“現状”，“影響”というクエリに関するサブテーマを示すような一般名詞のサブトピックが高スコアであった．それに対し，趣味に関するクエリに関しては，“豆知識”，“クイズ”などの知識に関する意図が低スコアである一方，学習に関するクエリでは低スコアであった“動画像”が高スコアであった．これは，遊ぶときに動画像を見るという連想がしやすい一方，学習のためにWeb検索で動画像を調べるとは連想しにくいためであると考えられる．しかし，動画像は学習の際の理解の支援になるので，クエリに関す

表 4.8: 低スコアな意図とサブトピック例（意図の例(サブトピック例:取得元ソース)）

地球温暖化	沖縄(学習)	沖縄(趣味)	ポケモン
動画像 (アニメーション:kidKey)	レストラン (居酒屋:genKey)	クイズ (クイズ:KidKey)	豆知識 (開発:wiki)
リンク集 (リンク集:kidCat)	サイト(総合情報サイト:genKey)	データ (人口:wiki)	スポット(ナガシマスパーランド:genKey,kidKey)
議論 (議論:wiki)	ニュース (ニュース:kidCat)	豆知識 (県名の由来:wiki)	地名 (イスラム諸国:wiki)

るサブトピックという扱いではなく，情報の形式という扱いで別に提示すべきであると考えられる．また，全クエリに共通する特徴として，“スポット”，“団体”，“地名”などの固有名詞が多い意図に関しては，子供にとって連想しやすい単語であるかかどうかが重要であると考えられる．例えば，“ナガシマスパーランド”と“イスラム諸国”は，ポケモンから想像しにくいと考えられるため，低スコアである．これらの結果より，学習と趣味によって必要とする意図が異なる点を考慮することが必要である．

表 4.5 と表 4.7，4.8 を比較すると，各ソースで抽出したサブトピックの表現する代表的な意図が，実際に子供にとって高スコアであるのは，“影響”，“スポット”，“気候”のみである．また，表 4.7 の取得元ソースに注目すると，高スコアな意図を表現するサブトピックは幅広いソースから取得されたものである．この結果より，子供にとって重要な意図を示すサブトピックを幅広く抽出するには，特定のソースを利用するだけでは不十分であることが分かる．

4. 学年ごとのスコアの比較

子供向けサブトピックの重要性を定量化する際に，子供のプロフィールを考慮する必要があるのかを明らかにするため，学年ごとのスコアの違いや特徴について述べる．小学校低学年については，被験者が 1 名であったため，除外する．

各学年でクエリやソースの違いとスコアの相関が異なるのかを調査するため，学年ごとの各クエリのスコアの平均値，学年ごとの各ソースのスコアの平均値を算出した．表 4.9 にクエリを検索目的にまとめたスコアの平均値，表 4.10 にソースごとのスコアの平均値を示

表 4.9: 各学年のクエリの検索目的ごとの平均スコア

	学年	学習	趣味
分かりやすさ	中1〜3	0.434	0.396
	小5〜6	0.394	0.484
	小3〜4	0.323	0.400
調べたさ	中1〜3	0.690	0.577
	小5〜6	0.514	0.487
	小3〜4	0.414	0.396
役立つかどうか	中1〜3	0.705	0.596
	小5〜6	0.507	0.455
	小3〜4	0.471	0.420

す．それぞれの学年のスコアについて，クエリとソースにおける2要因の分散分析を行ったところ，クエリについては，全ての学年で分かりやすさ，調べたさ，役立つかどうかの全ての観点で有意であった．一方，ソースについては，全ての学年の分かりやすさ，中1〜3と小5〜6の役立つかどうかの面で有意だった．つまり，クエリにおける趣味か学習という検索目的の違いは，どの学年の子供にとっても重要な指標であることが分かる．また，ソースによる違いは，どの学年の子供にとっても親しみやすさを考慮するときには重要であるが，単純に調べたいかどうかという面からはあまり重要ではない．役に立ちそうかどうかという面では，学年が上がればソースによる違いが重要になると考えられる．

表4.9に示すクエリに関する全体的な傾向として，学年が上がるにつれて，全体的なスコアが高くなった．ただし，趣味に関するクエリの分かりやすさについてのみ，小5〜6のスコアが一番高く，嗜好の影響が強く出ていることが分かる．具体的には，クエリ“ポケモン”について，小5〜6の男子が一番親しみやすかったことが原因であると考えられる．

また，表4.10に示すソースごとの結果の特徴的な点として，小3〜4の子供にとっては，分かりやすさではkidRelやgenRelのスコアが高い一方，調べたさや役立つかどうかでは全てのソースでスコアにあまり差がなかった．これは，小3〜4の子供にとってサブトピックを利用して検索を便利にするというイメージが乏しかったことが一つの原因と考えられる．つまり，幼い子供に対しては，単純にクエリとサブトピックの並びを提示するだけで

表 4.10: 各学年のソースごとの平均スコア

	学年	wiki	genRel	genKey	kidRel	kidKey	kidCat
分かりやすさ	中 1〜3	0.351	0.520	0.449	0.806	0.392	0.399
	小 5〜6	0.410	0.538	0.420	0.728	0.417	0.431
	小 3〜4	0.311	0.446	0.328	0.532	0.367	0.375
調べたさ	中 1〜3	0.650	0.693	0.631	0.817	0.623	0.647
	小 5〜6	0.517	0.557	0.479	0.640	0.499	0.487
	小 3〜4	0.395	0.444	0.414	0.419	0.416	0.361
役立つかどうか	中 1〜3	0.659	0.721	0.656	0.811	0.641	0.655
	小 5〜6	0.504	0.519	0.462	0.707	0.478	0.467
	小 3〜4	0.445	0.490	0.453	0.494	0.447	0.424

表 4.11: サブトピックの要素と各学年のスコアの相関係数

	学年	intent	difficulty	abstraction	kidHit	genHit
分かりやすさ	中 1〜3	0.383	−0.113	0.188	0.297	0.238
	小 5〜6	0.438	−0.211	0.239	0.258	0.322
	小 3〜4	0.374	−0.202	0.241	0.174	0.224
調べたさ	中 1〜3	0.372	−0.085	0.080	0.133	0.151
	小 5〜6	0.329	−0.104	0.097	0.159	0.172
	小 3〜4	0.356	−0.105	0.120	0.198	0.111
役立つかどうか	中 1〜3	0.378	−0.001	0.111	0.163	0.178
	小 5〜6	0.344	−0.059	0.178	0.214	0.149
	小 3〜4	0.343	−0.086	0.190	0.138	0.163

は不十分であると考えられる．一方，小 5〜6，中 1〜3 の子供にとっては，kidRel が一番高スコアであり，共通して有効なソースであることが分かるため，関連検索語を中心に他のソースからサブトピックを取得する方法が有効であると考えられる．

さらに，サブトピックの要素とサブトピックのスコアの関係が学年で変化するのかにつ

いて調査するために，学年ごとのスコアとサブトピックの要素の相関係数を算出した（表4.11）．この結果から，intentについては全学年で分かりやすさ，調べたさ，役立つかどうかの全ての側面で，弱から中程度の正の相関があることが分かる．difficultyについては，中1〜3では無相関であるが，小5〜6，小3〜4の分かりやすさとは弱い負の相関が見られた．abstractionについても，中1〜3では無相関であるが，小5〜6，小3〜4の分かりやすさとは弱い正の相関が見られた．kidHitとgenHitについては，全学年の分かりやすさで弱い正の相関が見られた．これらの結果より，スコアと相関する要素は学年ごとに異なることが分かる．具体的には，意図の違いや検索ヒット数は全学年で考慮しなければならず，難易度や抽象度は小学生以下の子供にとってはある程度考慮すべきであることが分かる．

5. 抽出サブトピックと入力サブトピックとの比較

本調査では，子供がどのようなサブトピックを連想できるのか，サブトピックの提示は必要なのかを調べるために，実験中に子供に自身が連想可能なサブトピックを入力してもらった．ここでは，それらのサブトピックの数や種類についての結果を述べ，考察を行う．

子供49人に入力してもらったサブトピックは11クエリで750個であった．つまり，子供が各クエリで連想できるサブトピックは平均1.74個という結果となった．学年別では，低学年で0個，中学年で0.64個，高学年で1.24個，中学生で1.50個であり，学年が低くなるほどサブトピックを連想することが難しくなることが分かる．これらのサブトピックの表記の揺れなどをまとめた結果，入力サブトピックの数は270個になった．入力してもらったサブトピックにどのような形式のものがあるのかを調査したところ，実際にWeb検索を行う際にはサブトピックとして直接利用できないものが多く含まれていた．そこで，子供に入力してもらったサブトピックを3人の評価者（大人）でアノテーションを行い，表4.12に示す，沖縄の“いいもの”のような曖昧なサブトピック，トマトの“赤い”，“おいしい”のような単に性質的な事実を述べているだけのサブトピック，地球温暖化の“どうして起こる？”という質問形式のサブトピック，実際に検索に利用できるサブトピック（‘一般’と記す）の4種類に分類した．3人の評価者間のFleiss's Kappaは0.831と高い一致度となった．検索に直接利用できないサブトピックは，全体の子供による入力サブトピックの24.6%を占めており，学年が低くなるほどそのようなサブトピックを入力する傾向が強かった．

また，今回実験で使用した様々なソースから抽出したサブトピックが，子供に入力してもらったサブトピックをどのくらいカバーしているのかを調べるために，入力サブトピッ

表 4.12: 入力サブトピックの形式ごとの占める割合

	曖昧	事実	質問	一般
全体	0.034	0.133	0.079	0.754
中 1〜3	0.019	0.087	0.025	0.869
小 5〜6	0.034	0.198	0.107	0.661
小 3〜4	0.087	0.131	0.174	0.608

クに対する抽出サブトピックの再現率を算出した．入力サブトピックは270個，抽出サブトピックと入力サブトピックで一致したサブトピックは98個であり，再現率は0.363となり，子供が連想しやすい表現のサブトピックを，ソースから直接抽出することは難しいことが分かる．逆に，様々なソースから抽出したサブトピックのうち，高スコアなサブトピックを子供がどのくらい連想できるのかを調べるために，分かりやすさ，調べたさ，役立つかについてのサブトピックのスコアの和が1.5より大きいのものを高スコアな抽出サブトピックとして，高スコアな抽出サブトピックに対する入力サブトピックの再現率を算出した．高スコアな抽出サブトピックは451個，高スコアな抽出サブトピックと入力サブトピックで一致したサブトピックは78個であり，再現率は0.172となり，ソースから抽出できる高スコアなサブトピックを子供が連想することは難しいことが分かる．このように，子供が自力で連想できるサブトピックは，表4.12の結果のように，不十分なものも多く，子供が簡単には連想できないが実は興味をひき役に立つと感じる抽出サブトピックを提示することで支援を行えると考えられる．

以上の結果より，子供がクエリのみから連想できるサブトピックは限られており，Web検索を支援する際に有用となるものは少なく，子供向けサブトピックの作成を支援することが重要であると考えられる．また，子供向けサブトピックを子供がクエリと結び付けやすい実体験に基づいた事実や質問という形式に変換して提示するのも効果的であると考えられる．

4.4 子供向けサブトピックランキング手法

上記の調査結果に基づき，子供向けサブトピックランキング手法について検討する．この手法を用いることで，多くのサブトピック候補の中から子供にとって分かりやすさ，調べたさ，役立つかどうかといった観点で重要度の高いサブトピックを優先的に提示することが可能となる．

4.4.1 手法の設計

子供向けサブトピックランキング手法では，複数のソースからサブトピック候補を抽出し，抽出ソースの種類に加えて，クエリの検索目的，サブトピックの要素を考慮して，サブトピックのスコアを算出し，ランキングを行う．具体的には，以下の手順で子供向けにサブトピックのランキングを行う．括弧内は4.3.2項に示すどの知見に対応する指標なのかを示している．

1. 複数のソースからサブトピック候補を抽出する．
   - Yahoo! JAPANにおけるクエリの関連検索語
   - Yahoo! JAPANにおけるクエリの検索結果上位100件のスニペットに出現するキーフレーズ
   - Yahoo!きっずにおけるクエリの関連検索語
   - Yahoo!きっずにおけるクエリの検索結果上位100件のスニペットに出現するキーフレーズ
   - Yahoo!きっずにおけるクエリの検索結果上位100件の登録カテゴリ
   - クエリをタイトルとするWikipediaの記事内の項目タイトル
2. 複数の指標に基づき各サブトピックのスコアを算出する．
   - クエリに関する指標（知見1）
     - クエリの検索目的：学習もしくは趣味.
   - 抽出ソースに関する指標（知見2）
     - genRel：Yahoo! JAPANの関連検索語かどうか.

- genKey : Yahoo! JAPAN の検索結果のスニペットに出現するキーフレーズかどうか．キーフレーズの重要度．
- kidRel : Yahoo!きっずの関連検索語かどうか．
- kidKey : Yahoo!きっずの検索結果のスニペットに出現するキーフレーズかどうか．キーフレーズの重要度．
- kidCat : Yahoo!きっずの検索結果中の登録カテゴリかどうか．
- wiki : Wikipedia の項目タイトルかどうか．

● サブトピックの要素に関する指標（知見 3）

- difficulty : 難易度．漢字やアルファベットの文字数．
- abstraction : 抽象度．一般名詞か固有名詞か．
- intent : 意図．あらかじめ定義した教科カテゴリと属性
- hit: Yahoo! JAPAN，Yahoo!きっずでの検索ヒット数．

3. スコアの降順にランキングする．

抽出ソースに関する指標の genKey，kidKey については有無だけでなく，Yahoo!キーフレーズ API で取得したキーフレーズの重要度を指標の値として用いる．サブトピックの要素に関する指標については，どのように各指標をスコア化するかは検討しなければならないが，本研究では，単純な方法でサブトピックの各要素を指標として用いたときに，どの程度効果があるのかを確認するため，難易度については漢字とアルファベット数，抽象度については一般名詞か固有名詞かを用いた．意図については，自動で決定することは難しいため，4.3 節で述べた調査の結果，子供にとってのサブトピックの重要度に影響を与えると予想される意図の一種となり得る教科カテゴリと属性を用いた．教科カテゴリは Yahoo!きっずのカテゴリである科学，数学，社会，生活，言語，総合学習，その他を用い，Yahoo!きっずの各カテゴリの登録サイトに出現する重要語を抽出して，各カテゴリの辞書を作成した．そして，サブトピックがどのカテゴリの辞書に登録された重要語を含むかで該当カテゴリを決定した．属性については，組織，人名，地名，サイト，時，情報タイプのあらかじめ定義したものから選択して用いた．

4.4.2 評価

データセット

データセットは4.3.1項で述べた子供49人に評価してもらった11種類のクエリに関するサブトピック1,100個とした．分かりやすさ，調べたさ，役立つかどうかのそれぞれの観点で子供に評価してもらったスコアの降順に並べたサブトピックのランキングを，それぞれの観点における正解ランキングとした．

評価指標

ランキングの精度を測る代表的な指標として，MAP値 [10] とNDCG値を用いた．

MAP(Mean Average Precision) は高スコアなサブトピックを網羅している割合を示す指標で，式(4.2) より求める．

$$AP_q = \frac{1}{k}\left(\sum_k rel_i\right) \tag{4.1}$$

$$MAP = \frac{1}{n}\left(\sum_n AP_i\right) \tag{4.2}$$

rel_i はランキング i 番目のサブトピックのデータセットで定義されたスコア，AP_q はクエリ q に関する Average Precision（式(4.1) より求める），k はクエリ数，n は全サブトピック数である．

一方，クエリ q に対するNDCGは，スコアの高い順に並べた理想的な結果とのずれを表す指標で，式(4.3) より求める．

$$NDCG_q = \frac{1}{IDCG_q}\left(rel_1 + \sum_{i=2}^{l}\frac{rel_i}{log_2 i}\right) \tag{4.3}$$

rel_i はランキング i 番目のサブトピックのデータセットで定義されたスコア，l はランキングの検索数であり，上位1件から20件で評価を行った．$IDCG_q$ はクエリ q に対する $NDCG_q$ の理想値，つまりデータセットで定義された理想ランキングの $NDCG$ 値である．

評価結果

評価では，ソースを組み合わせる効果，サブトピックの要素を組み合わせる効果を調査するため，ソース単独を用いたもの（genRel，genKey，kidRel，kidKey，kidCat，wiki），

表 4.13: 指標の各組合せによるランキングの MAP 値

	wiki	genRel	genKey	kidRel	kidKey	kidCat	Source	All
分かりやすさ	0.116	0.300	0.033	0.091	0.116	0.014	**0.389**	**0.422**
調べたさ	0.033	0.035	0.016	0.090	0.026	0.092	**0.511**	**0.524**
役立つかどうか	0.176	0.314	0.043	0.090	0.106	0.022	**0.512**	**0.530**

ソース全てを用いたもの（Source），全ての指標を用いたもの（All）の 8 種類のケース（手法）でランキングした結果を比較した．

指標の組合せは Ranking SVM [36] を用いて決定した．具体的には，カーネルは線形カーネルを用い，データセット中の 1 種類のクエリをテストデータ，残りの 10 種類のクエリをトレーニングデータとし，Ranking SVM を用いてランキング関数を学習した．学習したランキング関数によって，テストデータのサブトピックのランキングを作成し，NDCG 値と AP 値を算出した．全てのクエリを順にテストデータとし，最終的に，全 11 種類のクエリの平均 NDCG 値と MAP 値を算出した．

各手法における MAP 値を表 4.13 に示す．有意水準 0.05 で Bonferroni 法で多重比較を行ったところ，MAP 値については，全ての質問に対して，ソース単独よりも Source と All が有意に高精度であった．具体的には，All については，分かりやすさで 12.2%から 40.8%，調べたさで 43.2%から 49.8%，役に立つかどうかで 21.6%から 50.8%，ソース単独よりも MAP 値が向上した．また，Source については，分かりやすさで 8.9%から 37.5%，調べたさで 41.9%から 49.5%，役に立つかどうかで 19.6%から 49.0%，ソース単独よりも MAP 値が向上した．これは，ソース単独では取得できるサブトピックに限界があるためである．Source と All の MAP 値については有意差は見られなかったが，All の方が安定して高精度であった．具体的には，All については，分かりやすさで 3.3%，調べたさで 1.3%，役に立つかどうかで 1.8%，Source よりも MAP 値が向上した．

各手法における NDCG 値を図 4.2 に示す．横軸は，式 (4.3) における上位何番目までのサブトピックを検索するかという検索数 l を示す．NDCG 値については，単独ソースの結果では，分かりやすさや役に立つかどうかで，ランキング上位での genRel や kidRel は比較的高精度であり，分かりやすさのみでは genKey も比較的高精度であった．しかし，ランキング 10 位を超えると途端に精度が下がったり，質問ごとに偏りがあったりするため，単独

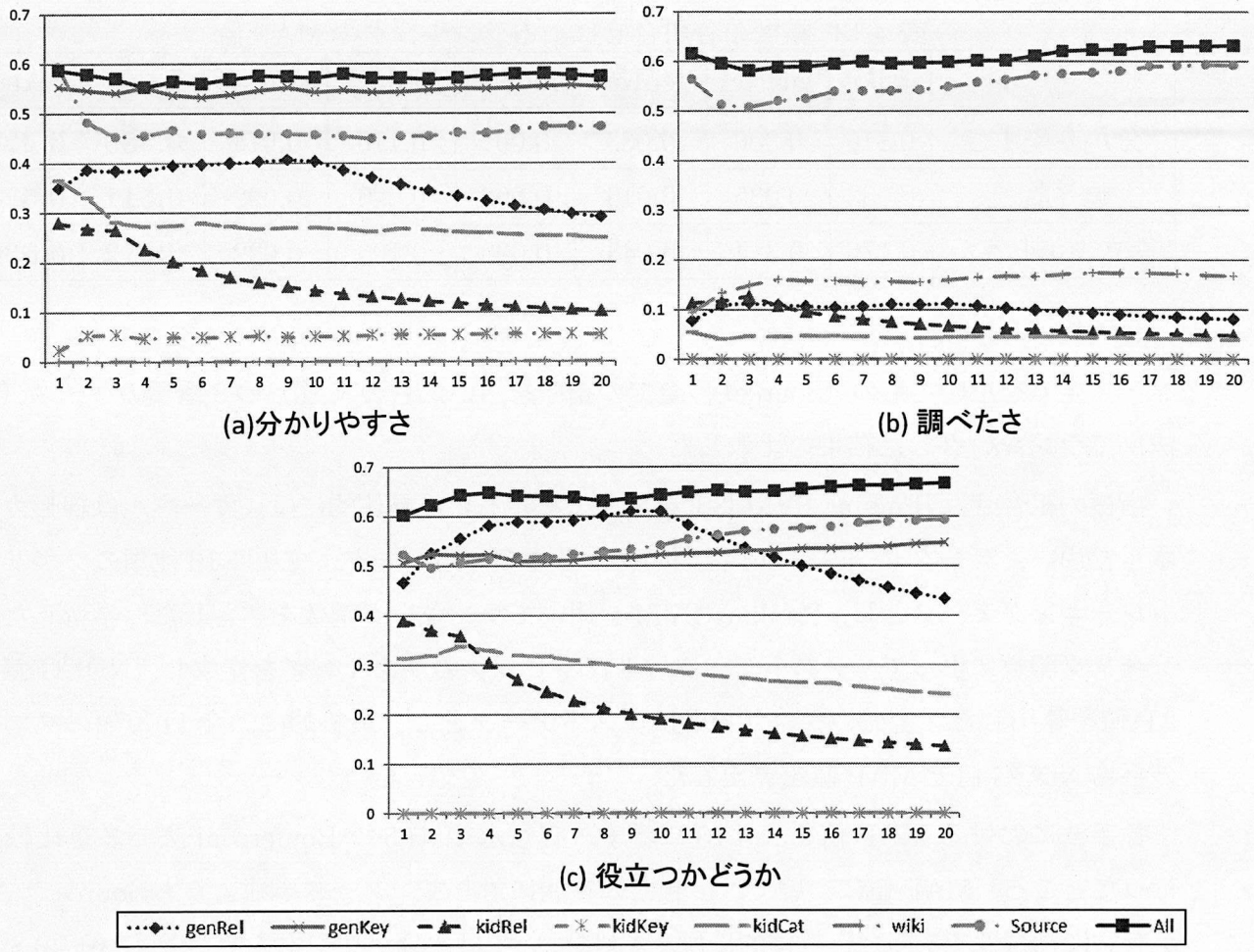


図 4.2: 指標の各組合せによるランキングのNDCG値

ソースでの精度は不安定であるといえる．それに対して，SourceやAllは安定して高精度であり，さらにAllはSourceよりも高精度である．有意水準0.05でBonferroni法で多重比較を行ったところ，全ての質問で，ほぼ全ての点に対して，AllのNDCG値は単独ソースのNDCG値だけでなく，SourceのNDCG値を有意に上回っていた．具体的には，Allは，分かりやすさで0%から12.4%，調べたさで3.8%から8.2%，役に立つかどうかで7.3%から13.6%，SourceのNDCG値を上回っていた．これらの結果より，単独ソースよりもソースを複数用いる方が幅広く安定して高精度にサブトピックを取得できること，ソースだけでなく，サブトピックの要素を考慮することで，より有効なサブトピックを上位にランクできることを確認した．

学年ごとの差について調べるために，学習数は少なくなってしまうが，学年別のデータ

表 4.14: 各学年ごとの All によるランキングの MAP 値

	全学年	中 1-3	小 5-6	小 3-4
分かりやすさ	0.422	0.416	0.426	0.362
調べたさ	0.524	**0.635**	0.508	0.392
役立つかどうか	0.530	**0.626**	0.476	0.451

セットを作成し，学年ごとの正解ランキングによって学習したランキング関数を利用して，サブトピックをランキングした際の精度を算出した．学年ごとの MAP 値を表 4.14，NDCG 値を図 4.3 に示す．調べたさと役立つかどうかに関して，中学生では，MAP 値，ランク上位での NDCG 値が全学年の値よりも向上した．しかし，全体的には，学年別にしたことによる大きな精度の向上は見られなかった．これは，学年ごとのデータセットでは，各サブトピックを評価した子供の人数が少なく，十分な学習数を確保できなかったためと考えられる．そのため，学年ごとの違いをより詳細に調査するため，より大規模な人数にサブトピックの評価をしてもらう必要がある．

4.5 むすび

本章では，子供にとって重要なサブトピックを子供に優先的に提供するため，1) 子供向けサブトピックの調査，および，2) 子供向けサブトピックランキング手法の検討について取り組んだ．子供向けサブトピックの調査に関しては，小中学生 49 名に，一般向け検索エンジン，子供向け検索エンジン，Wikipedia という複数のソースから取得したサブトピックを分かりやすさ，調べたさ，役立つかどうかの 3 つの観点からスコア付けしてもらうことで行った．そして，高スコアのサブトピックを取得できるクエリやソースの特徴，難易度などのサブトピックの要素とスコアの関係，学年ごとのスコアの比較という観点から，スコア付けしてもらった結果の調査を行った．その後，調査の結果に基づいて，子供向けサブトピックランキング手法について検討した．子供向けサブトピックのランキングは，取得ソースの種類と難易度などのサブトピックの要素に関する指標をもとに，サブトピックの子供向け度合いを数値化することで行う．子供にスコア付けしてもらったデータセットを用いて，子供向けサブトピックランキング手法で順位付けしたサブトピックのランキン

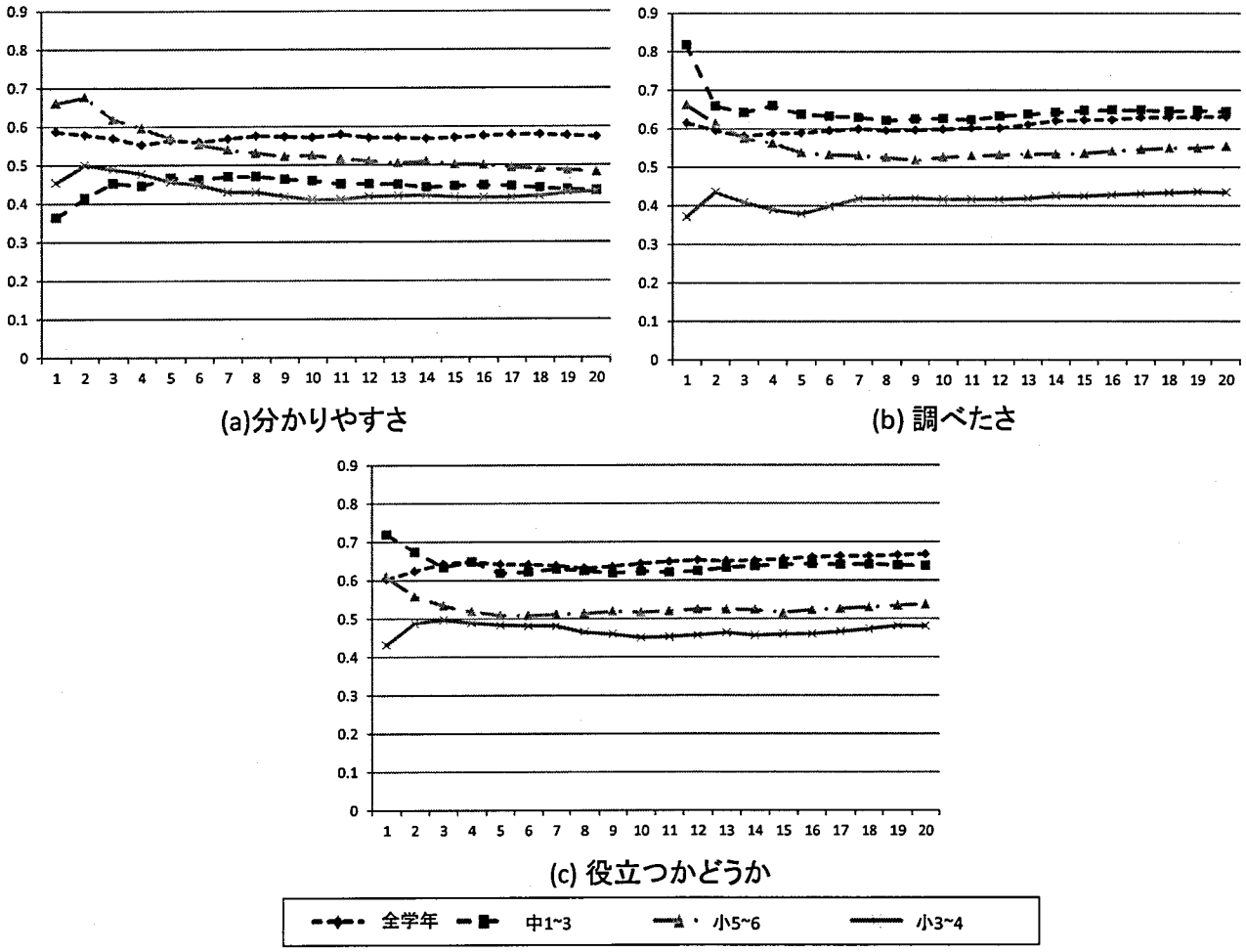


図 4.3: 各学年ごとの All によるランキングの NDCG 値

グの精度を確認したところ，提案手法では，サブトピックの抽出元ソースの種類と，サブトピック自身の難易度，教科カテゴリ，検索ヒット数といったシンプルな指標を考慮するだけで，ソースを単独で用いたり，抽出元ソースの種類のみを考慮してサブトピックをランキングするよりも，子供にとって子供向け度合いの高いサブトピックを多く上位にランクできたことを確認した．

第5章

結論

5.1 本論文のまとめ

本論文では，子供が興味を持ち，分かりやすくWeb上の情報を取得できるように，子供のWeb閲覧・Web検索の支援を行う機構について検討した．近年，小学生を始めとする子供がインターネットを利用し，Web上の情報を閲覧，検索することが一般的となってきており，このような子供の特徴を考慮したWeb閲覧・Web検索支援技術の確立は，非常に重要な課題である．

まず第1章で，これまでに行われてきたWeb閲覧・Web検索に関連する子供の特徴を調査した研究，子供向けに考案されたWebシステムに関する研究について整理し，小学生を始めとする子供に特化したWeb閲覧・Web検索の支援を行うことの重要性について述べた．

第2章では，難解なページであっても，子供が興味を持って分かりやすく閲覧できるような子供向けWebブラウザの検討を行った．本ブラウザでは，子供の特徴を考慮したデザインとして泡メタファを用いた海中アニメーションによるWebページの提示を行う．具体的には，Webページをコンポーネントに分割し，コンポーネントごとに泡を生成して表示する．また，コンポーネントの詳細内容は子供が楽しく理解しやすいように絵本形式へ変換して提示する．これにより，子供にとって理解が難しい一般向けWebページでも，楽しくわかりやすくWeb閲覧することができる．評価実験の結果，比較対象として用いた漢字を平仮名に変換して提示するよりブラウザよりもページ内容の質問に対する正答率が20%向上したことから，子供にとって分かりやすい提示を実現できたと考える．また，Webページを元の形態のまま提示するブラウザよりも楽しいと答えた被験者が50%増加したこ

とから，一般のブラウザよりも画像を用いることで子供の興味をひく提示を実現できたと考える．

次に第3章では，Web検索を行う際に，第1章で述べた方法でWebページを変換せずとも，元々子供にとって十分興味をひき，分かりやすいような子供向けのWebページを簡単に取得できるよう，Web検索結果の子供向けリランク手法の検討を行った．具体的には，提案手法では，Webページの文章量や色の数などの構成に関する指標，文の長さや子供向け表現の数などの文章に関する指標をスコア化する．そして，各指標のスコアを組み合わせて各ページのスコアを決定し，スコアの降順に検索結果を並べ替える．評価実験の結果，子供向け表現を中心とした指標を組み合わせることで，子供にとっての“読みたさ”，“見た目の見やすさ”，“役に立つか”，“内容の分かりやすさ”の4つの側面において，本手法によるリランクは，Yahoo! JAPANおよび，Yahoo!きっずのランキングよりNDCG値が最大20%向上した．この結果から，提案手法により，子供向け度合いの高いページを上位にランクでき，子供が検索結果から分かりやすいページを容易に選択できるようになることが期待される．

第4章では，WebページやWeb検索結果という出力だけでなく，入力としてのWeb検索方法の支援を行うため，検索クエリに関する子供向けサブトピックのランキング手法の検討を行った．具体的には，子供にとって重要なサブトピックを子供に優先的に提供するため，1) 子供向けサブトピックの調査，および，2) 子供向けサブトピックランキング手法の検討について取り組んだ．子供向けサブトピックの調査に関しては，子供に，一般向け検索エンジン，子供向け検索エンジン，Wikipediaという複数のソースから取得したサブトピックを“調べたさ”，“役に立つか”，“分かりやすさ”の3つの観点からスコア付けしてもらうことで行った．そして，高スコアのサブトピックを取得できるクエリやソースの特徴，難易度などのサブトピックの要素とスコアの関係，学年ごとのスコアの比較という観点から，スコア付けしてもらった結果の調査を行った．その後，調査の結果に基づいて，子供向けサブトピックランキング手法を提案した．子供向けサブトピックのランキングは，取得ソースの種類と難易度などのサブトピックの要素に関する素性をもとに，サブトピックの子供向け度合いを数値化することで行う．評価実験の結果，提案手法は，ソースを単独で用いたり，ソースの種類を考慮してサブトピックをランキングするよりも，子供にとって子供向け度合いの高いサブトピックを多く上位にランクできたことを確認した．

本論文の成果により，これまで，Web閲覧・Web検索を十分に快適に利用することが困

難であった子供にとって，興味をひき，分かりやすい形式で，Webページの閲覧やWeb検索を行うことが可能となる基盤を構築できた．

5.2 今後の課題

第2章の研究課題では，海中アニメーション画面，詳細内容表示画面とも，さらに，子供にとって分かりやすく興味を引く方法について検討する必要がある．例えば，今回はどのようなWebページであっても，海中アニメーションと絵本による形式に変換したが，Webページの内容が属するカテゴリによって，アニメーションのタイプや絵本画面での会話のタイプを変えることで，子供にとって，より興味をひき，分かりやすい表示形式になると考えられる．また，提案ブラウザを長期的に使用してもらい，“飽きないかどうか”についても調査する必要がある．

第3章の提案手法では，単純にテキスト量が少ないほど，画像量が多いほど，ページの子供向け度合いとしてのスコアが高くなるように設定した．しかし，実験の結果，指標ごとに学年に適した値があることが示唆された．そのため，テキスト量，画像量，文章の難易度などの年齢に応じた適切な値を調査し，指標のスコア化の方法をさらに改善する必要がある．さらに，提案手法の効果は，クエリや学年による差も大きく影響する結果となったため，クエリや学年の特徴に従った各指標の最適な組合せ方法を検討することも必要である．

第4章の研究課題では，子供目線での重要度のみを考慮していたので，子供にとって重要なサブトピックの中で，実際にWeb検索時に役立つサブトピックがどのようなものなのかをさらに調査する必要がある．また，今回使用した意図，抽象度，難易度，検索ヒット数というサブトピックの特徴の他にも有効な特徴がないのかをさらに検証し，サブトピックランキング手法を改善する必要がある．

一方，これまでの研究全般において，子供のWeb閲覧・Web検索の能力および特徴に，学年による差が大きいことが明らかとなった．子供は学年が上がるにつれて，学校で学ぶ内容を始めとし，著しく知識量が増える．そのため，本研究全体の課題として，学年ごとのWebページやサブトピックに対する嗜好などの特徴をより詳細に調査し，子供の学年ごとの違いを考慮した支援を行う必要がある．また，これまでの手法を組み合わせ，総合的に子供のWeb閲覧・Web検索の支援を行うシステムの開発を行う必要がある．さらに，これまでの研究では，子供の興味をひく情報を優先的に提供する方法に着目していたが，子

供が閲覧する情報を限定することは，逆に，意図しないものと巡り合う可能性が減ってしまい，子供がWeb上の情報に対してひ弱になるということも考えられる．そのため，年齢が上がるにつれて，限定する情報の幅を狭めたり，子供の興味をひく情報とそれ以外の情報のバランスを親や学校が調節できるような仕組みについても検討する必要がある．

謝辞

本研究全般に関して，懇切なる御指導と惜しみない御助言を頂きました大阪大学大学院情報科学研究科マルチメディア工学専攻 西尾章治郎教授に謹んで御礼申し上げます.

本研究を推進するにあたり，直接の御指導，御助言，御討論を頂きました大阪大学大学院情報科学研究科マルチメディア工学専攻 原隆浩准教授に衷心より感謝申し上げます.

本論文をまとめるにあたり，大変有益な御指導と御助言を多数賜りました大阪大学サイバーメディアセンター 下條真司教授，大阪大学大学院情報科学研究科マルチメディア工学専攻 細田耕教授に心より感謝申し上げます.

講義，学生生活を通じて，学問に取り組む姿勢をご教授頂きました大阪大学大学院情報科学研究科マルチメディア工学専攻 藤原融教授，薦田憲久教授に厚く感謝申し上げます.

本研究において，多大なる御助言，御協力，御支援を頂きました独立行政法人情報通信研究機構 寺西裕一博士，大阪大学大学院工学研究科 春本要准教授，神戸大学大学院工学研究科 寺田努准教授，大阪大学サイバーメディアセンター 義久智樹准教授，大阪大学大学院情報科学研究科マルチメディア工学専攻 神崎映光助教，東京大学知の構造化センター 中山浩太郎特任講師に深謝致します.

筆者の所属する研究グループにおいて，有益な御助言を頂いた，大阪大学大学院情報科学研究科マルチメディア工学専攻 白川真澄氏，宮本大樹氏，山本彩奈氏，大澤純氏，杉谷卓哉氏，本田博之氏，加藤諒氏，中村達哉氏に深く御礼申し上げます.

本研究を進める上で惜しみない御助言，御協力，研究活動を進めるにあたっての多大なる御支援を頂いた，株式会社 KDDI 研究所 服部元博士，上向俊晃博士，Microsoft Research Asia 酒井哲也博士，荒瀬由紀博士，東芝株式会社 伊藤雅弘博士，株式会社四国電力 大西健史氏，富士通株式会社 小牧大治郎博士，日本電信電話株式会社 足利えりか氏，トヨタ自動車株式会社 奥梓氏，みずほコーポレート銀行 鈴木晃祥氏に感謝の意を表します.

本研究を進めるにあたり，多くの御討論や御助言を頂きました大阪大学大学院情報科学

研究科マルチメディア工学専攻 西尾研究室の諸氏に心より感謝申し上げます.

最後に，研究生活を送る上で，暖かい御支援と多大なる御理解を頂いた両親を始めとする家族に心からの感謝と御礼を申し上げます.

参考文献

[1] Agrawal, R., Gollapudi, S., Halverson, A., and Ieong, S.: Diversifying Search Results, in *Proceedings of International Conference on Web Search and Data Mining (WSDM2009)*, pp. 5–14, 2009.

[2] Anderson, J.R.: Cognitive Psychology and Its Implications, *Freeman & Co*, 1980.

[3] Anscombe, F.J.: The Validity of Comparative Experiments, *Journal of the Royal Statistical Society*, Vol. 111, No. 3, pp. 181–211, 1948.

[4] 荒瀬由紀, 前川卓也, 原隆浩, 上向俊晃, 西尾章治郎: 携帯電話を用いた Web 閲覧のためのコンテンツ適応的提示システム, 情報処理学会論文誌, Vol. 47, No. 12, pp. 3149–3164, 2006.

[5] Azzopardi, L., Glassey, R., Mounia, L., Polajnar, T., and Ruthven, I.: PuppyIR: Designing an Open Source Framework for Interactive Information Services for Children, in *Proceedings of Workshop on Human-Computer Interaction and Information Retrieval (HCIR 2009)*, 2009.

[6] Bilal, D.: Children's Use of the Yahooligans! Web Search Engine: I. Cognitive, Physical, and Affective Behaviors on Fact-based Search Tasks, *Journal of the American Society for Information Science*, Vol. 51, No. 7, pp, 646–665 (2000).

[7] Bilal, D.: Children's use of the Yahooligans! Web search engine: II. Cognitive and physical behaviors on research tasks, *Journal of the American Society for Information Science*, Vol. 52, No. 2, pp. 118–136, 2001.

[8] Bilal, D.: Children's Use of the Yahooligans! Web Search Engine III. Cognitive and Physical Behaviors on Fully Self-Generated Search Tasks, *Journal of the American Society for Information Science and Technology*, Vol. 53, No. 13, pp. 1170–1183, 2002.

[9] Bilal, D. and Kirby, J.: Differences and Similarities in Information Seeking: Children and Adults as Web Users, *Information Processing and Management*, Vol. 38, No. 5, pp. 649–670, 2002.

[10] Buckley, C. and Voorhees, E.M.: Evaluating Evaluation Measure Stability, in *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2000)*, pp. 33–40, 2000.

[11] De Belder, J. and Moens, M.-F.: Text Simplification for Children, in *Proceedings of Workshop on Accessible Search Systems of International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2010)*, pp. 19–26, 2010.

[12] De Belder, J. and Moens, M.-F.: A Dataset for the Evaluation of Lexical Simplification, in *Proceedings of International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2012)*, pp. 426–437, 2012.

[13] Dang, V., Xue, X., and Croft, B.: Inferring Query Aspects from Reformulations using Clustering, in *Proceedings of ACM International Conference on Information and Knowledge Management (CIKM 2011)*, pp. 2117–2120, 2011.

[14] Dou, Z., Hu, S., Chen, K., Song, R., and Wen, J.-R.: Multi-Dimensional Search Result Diversification, in *Proceedings of ACM International Conference on Web Search and Data Mining (WSDM 2011)*, pp. 475–484, 2011.

[15] Druin, A., Foss, E., Hutchinson, H., Golub, E., and Hatley, L.: Children's Roles using Keyword Search Interfaces at Home, in *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI 2010)*, pp. 413–422, 2010.

[16] Druin, A., Foss, E., Hatley, L., Golub, E., Guha, M.L., Fails, J., and Hutchinson, H.: How Children Search the Internet with Keyword Interfaces, in *Proceedings of*

International Conference on Interaction Design and Children (IDC 2009), pp. 89–96, 2009.

[17] Duarte Torres, S., Hiemstra, D. and Serdyukov, P.: An Analysis of Queries Intended to Search Information for Children, in *Proceedings of Symposium on Information Interaction in Context (IIiX 2010)*, pp. 393–402, 2010.

[18] Duarte Torres, S., Hiemstra, D. and Serdyukov, P.: Query Log Analysis in the Context of Retrieval of Information for Children, in *Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2010)*, pp. 847–848, 2010.

[19] Duarte Torres, S. and Weber, I.: What and How Children Search on the Web, in *Proceedings of ACM International Conference on Information and Knowledge Management (CIKM 2011)*, pp. 393–402, 2011.

[20] Dunn, O.J.: Multiple Comparisons Among Means, *Journal of the American Statistical Association*, Vol. 56, No. 293, No. 52–64, 1961.

[21] Dunnett, C.W.: A Multiple Comparisons Procedure for Comparing Several Treatments with a Control, *Journal of the American Statistical Association*, Vol. 50, No. 272, pp. 1096–1121, 1955.

[22] Dunnett, C.W..: New Tables for Multiple Comparisons with a Control, *Biometrics*, Vol. 20, No. 3, pp. 482–491, 1964.

[23] Eickhoff, C., Serdyukov, P., and Vries, A.P.: A Combined Topical/non-topical Approach to Identifying Web Sites for Children, in *Proceedings of ACM International Conference on Web Search and Data Mining (WSDM 2011)*, pp. 505–514, 2011.

[24] Eickhoff, C., Serdyukov, P., and Vries, A.P.: Web Page Classification on Child Suitability, in *Proceedings of ACM International Conference on Information and Knowledge Management (CIKM 2010)*, pp. 1425–1428, 2010.

[25] Elliot, D., Glassey, R., Polajnar, T., and Azzopardi, L.: Filtering and Finding Information for Children, in *Proceedings of International ACM Conference on the Research and Development of Information Retrieval (SIGIR 2010)*, p. 702, 2010.

[26] Fleiss, J.L.: Measuring Nominal Scale Agreement among Many Raters, *Psychological Bulletin*, Vol. 76, No. 5 pp. 378–382, 1971.

[27] Foss, E., Druin, A., Brewer, R., Lo, P., Sanchez, L., E., Golub, and Hutchinson, H.: Children's Search Roles at Home: Implications for Designers, Researchers, Educators, and Parents, *Journal of the American Society for Information Science and Technology*, Vol. 63, No. 3, pp. 558–573, 2012.

[28] Gerzog, E.H. and Haugland, S.W.: Web Sites Provide Unique Learning Opportunities for Young Children, *Early Childhood Education Journal*, Vol. 27, No. 2, pp. 209–114, 1999.

[29] Glassey, R., Elliott, D., Polajnar, T., and Azzopardi, L.: Interaction-based Information Filtering for Children, in *Proceedings of Symposium on Information Interaction in Context (IIiX 2010)*, pp. 329–334, 2010.

[30] Glassey, R., Polajnar, T. and Azzopardi, L.: PuppyIR Unleashed: A Framework for Building Child-Oriented Information Services, In *Proceedings of Dutch-Belgian Information Retrieval Workshop (DIR 2011)*, 2011.

[31] Gossen, T., Low, T., and Nürnberger, A.: What are the Real Differences of Children's and Adults' Web Search?, in *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2011)*, pp. 1115–1116, 2011.

[32] Gyllstrom, K. and Moens, M.-F.: A Picture is Worth a Thousand Search Results: Finding Child-oriented Multimedia Results with CollAge, in *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2010)*, pp. 731–732, 2010.

[33] Gyllstrom, K. and Moens, M.-F.: Clash of the Typings: Finding Controversies and Children's Topics within Queries, in *Proceedings of European Conference on Advances in Information Retrieval (ECIR 2011)*, pp. 80–91, 2011.

[34] Gyllstrom, K. and Moens, M.-F.: Wisdom of the Ages: toward Delivering the Children's Web with the Link-based Agerank Algorithm, in *Proceedings of ACM International Conference on Information and Knowledge Management (CIKM 2010)*, pp. 159–168, 2010.

[35] He, J., Hollink, V., and Vries, A.: Combining Implicit and Explicit Topic Representations for Result Diversification, in *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2012)*, pp. 235–244, 2012.

[36] Herbrich, R., Graepel, T., and Obermayer, K.: Large Margin Rank Boundaries for Ordinal Regression, *Advances in Large Margin Classifiers*, MIT Press, pp. 115–132, 2000.

[37] Hu, J., Wang, G., Lochovsky, F., Sun, J.-T., and Chen, Z.: Understanding User's Query Intent with Wikipedia, in *Proceedings of International Conference on World Wide Web (WWW 2009)*, pp. 471–480, 2009.

[38] Hu, Y., Qian, Y., Li, H., Jiang, D., Pei, J., and Zheng, Q.: Mining Query Subtopics from Search Log Data, in *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2012)*, pp. 305–314, 2012.

[39] 蓬莱博哉, 灘本明代, 田中克己: 理解しやすさとユーモアを考慮した Web コンテンツの対話番組変換, 日本データベース学会 Letters, Vol. 2, No. 2, pp. 29–32, 2003.

[40] Hourcade, J. and Bederson, B.: Differences in Pointing Task Performance between Preschool Children and Adults Using Mice, *ACM Transactions on Computer-Human Interaction*, pp. 357–386, 2004.

[41] Hu, J. and Bagga, A.: Identifying Story and Preview Images in News Web Pages, in *Proceedings of Conference on Document Analysis and Recognition (ICDAR 2003)*,

pp. 640–644, 2003.

[42] Hutchinson, H.B., Bederson, B.B., and Druin, A.: The Evolution of the International Children's Digital Library Searching and Browsing Interface, in *Proceedings of Conference on Interaction design and children (IDC 2006)*, pp. 105–112, 2006.

[43] 稲垣宣生, 山根芳知, 吉田光雄: 統計学入門, 裳華房, 1992.

[44] Isabelle G., André E.: An Introduction to Variable and Feature Selection, *Journal of Machine Learning Research*, Vol. 3, pp. 1157–1182, 2003.

[45] Iwata, M., Arase, Y., Hara, T., and Nishio, S.: A Children-Oriented Re-ranking Method for Web Search Engines, in *Proceedings of International Conference on Web Information System Engineering (WISE 2010)*, pp. 225–239, 2010.

[46] 岩田麻佑, 荒瀬由紀, 原 隆浩, 西尾章治郎: 泡メタファーを用いた子供向け Web 提示手法について, 情報処理学会マルチメディア, 分散, 協調とモバイルシンポジウム (DICOMO 2008) 論文集, pp. 553–560, 2008.

[47] Iwata, M., Hara, T., and Nishio, S.: How Different are Preferences of Children and Adults on Web pages?, in *Proceedings of International Workshop on Advanced Distributed and Parallel Network Applications (ADPNA 2011)*, pp. 394–399, 2011.

[48] Iwata, M., Arase, Y., Hara, T., and Nishio, S.: Investigation of Children's Characteristics for Web Browsing, in *Proceedings of International Workshop on Web Information and Data Management (WIDM 2009)*, pp. 91–94, 2009.

[49] 岩田麻佑, 荒瀬由紀, 原 隆浩, 西尾章治郎: 子供による Web 検索のための検索結果リランク手法, 情報処理学会論文誌, Vol. 52, No. 3, pp. 1055–1068, 2011.

[50] Iwata, M., Arase, Y., Hara, T., and Nishio, S.: Web Browser for Children Using Bubble Metaphor, *International Journal of Web Information Systems*, Vol. 6, No. 1, pp. 55–73, 2010.

[51] 岩田麻佑, 荒瀬由紀, 原 隆浩, 西尾章治郎: Web ページの構成と文章に着目した Web 検索結果の子供向けリランク手法, 情報処理学会マルチメディア, 分散, 協調とモバイル (DICOMO 2010) シンポジウム論文集, pp. 1365-1372, 2010.

[52] 岩田麻佑, 小牧大治郎, 荒瀬由紀, 原 隆浩, 西尾章治郎: 泡メタファを用いた子供向けWebブラウザの評価, 情報処理学会マルチメディア, 分散, 協調とモバイルシンポジウム (DICOMO 2009) 論文集, pp. 1415–1423, 2009.

[53] 岩田麻佑, 原 隆浩, 西尾章治郎: 子供によるWeb検索支援のための検索クエリに関する子供向けサブトピックの調査, 情報アクセスシンポジウム 2012, pp. 23–30.

[54] 岩田麻佑, 原 隆浩, 西尾章治郎: Webページに対する子供と大人の嗜好に関する調査, 平成22年度情報処理学会関西支部支部大会講演論文集 (CD-ROM), 2010.

[55] Jarvelin, K. and Kekalainen, J.: Cumulated gain-based evaluation of IR techniques, *ACM Transactions on Information Systems*, Vol. 20, No. 4, pp. 422–446, 2002.

[56] Kammerer, Y. and Bohnacker, M.: Children's Web Search with Google: the Effectiveness of Natural Language Queries, in *Proceedings of International Conference on Interaction Design and Children (IDC 2012)*, pp. 184–187, 2012.

[57] 菊野春雄: 情報処理理論と幼児教育ー作業記憶容量と精微化の観点からー, エデュケア, Vol. 18, pp. 29–36, 1998.

[58] 菊地秀文, 赤堀侃司: 小学校情報教育における児童のWebブラウジングの特徴分析, 日本教育工学会論文誌, Vol. 27, No. 2, pp. 143–153, 2003.

[59] Kikuchi, H., Kato, H., and Akahori, K.: Analysis of Children's Web Browsing Process: ICT Education in Elementary Schools, in *Proceedings of International Conference on Computers in Education (ICCE 2002)*, pp. 253-254, 2002.

[60] Lingnau, A., Ruthven, I., Landoni, M., Van der Sluis, F.: Interactive Search Interfaces for Young Children? The PuppyIR Approach, in *Proceedings of IEEE International Conference on Advanced Learning Technologies (ICALT 2010)*, pp. 389–390, 2010.

[61] Maekawa, T., Hara, T., and Nishio, S.: Image Classification for Mobile Web Browsing, in *Proceedings of International Conference on World Wide Web (WWW 2006)*, pp. 43–52, 2006.

[62] 前川卓也, 原 隆浩, 西尾章治郎: モバイル端末のためのWebページ自動スクロール方式, 日本データベース学会Letters, Vol. 4, No. 2, pp. 29–32, 2005.

[63] Michael, O.: Sensory Evaluation of Food: Statistical Methods and Procedures, *CRC Press.* p. 487, 1986.

[64] Jochmann-Mannak, H., Huibers, T., Lentz, L., and Sanders, T.: Children Searching Information on the Internet: Performance on Children's Interfaces Compared to Google, in *Proceedings of Workshop on Accessible Search Systems of International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2010)*, pp. 27–35, 2010.

[65] 美馬秀樹, 尹 泰聖: 子供のためのウェブ情報検索支援システムの開発, 情報処理学会夏のプログラミング・シンポジウム報告集, pp. 17–23, 2003.

[66] 美馬秀樹, 尹 泰聖: 子供のためのウェブ情報検索支援システムの開発, 情報処理推進機構: 未踏ソフトウェア創造事業, 2004.

[67] Moore, P. and George, A.: Children as Information Seekers: The Cognitive Demands of Books and Library Systems, *School Library Media Quarterly*, Vol. 19, No. 3, pp. 161–168, 1991.

[68] 森由美子: 子供たちの異文化間コミュニケーション, 情報処理学会論文誌, Vol. 47, No. 3, pp. 276–282, 2006.

[69] Nakaoka, M., Shirota, Y., and Tanaka, K.: Web Information Retrieval Using Ontology for Children based on Their Lifestyles, in *Proceedings of International Conference on Data Engineering Workshops (ICDEW 2005)*, p. 1260, 2005.

[70] Nakayama, K., Hara, T., and Nishio, S.: Wikipedia Mining for an Association Web Thesaurus Construction, in *Proceedings of Conference on Web Information Systems Engineering (WISE 2007)*, pp. 322–334, 2007.

[71] Nielsen, J.: Usability of Websites for Children: Design Guidelines for Targeting Users Aged 3-12 Years, *Nielsen Norman Group Report*, 2009.

[72] Nielsen, J.: Teenagers on the Web: 61 Usability Guidelines for Creating Compelling Websites for Teens, *Nielsen Norman Group Report*, 2009.

[73] Nielsen, J.: Usability of Websites for Children: 70 Design Guidlines based on Usability Studies with Kids, *Nielsen Norman Group Report*, 2002.

[74] Nielsen, J.: Teenagers on the Web: 61 Usability Guidelines for Creating Compelling Websites for Teens, *Nielsen Norman Group Report*, 2005.

[75] 高村大也, 奥村 学: 言語処理のための機械学習入門, コロナ社, 2010.

[76] Pandey, S. and Punera, K.: Unsupervised Extraction of Template Structure in Web Search Queries, in *Proceedings of International Conference on World Wide Web (WWW 2012)*, pp. 409–418, 2012.

[77] Polajnar, T., Glassey, R., and Azzopardi, L.: Detection of News Feeds Items Appropriate for Children, in *Proceedings of European Conference on Advances in Information Retrieval (ECIR 2012)*, pp. 63–72, 2012.

[78] Polajnar, T. Glassey, R., and Azzopardi, L. : JuSe: a Picture Dictionary Query System for Children, in *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2011)*, pp. 1281–1282, 2011.

[79] Radlinski, F., Szummer, M., and Craswell, N.: Inferring Query Intent from Reformulations and Clicks, in *Proceedings of International World Wide Web Conference (WWW 2010)*, pp. 1171-1172, 2010.

[80] Santos, R.L.T., Macdonald, C., and Ounis, I.: Exploiting Query Reformulations for Web Search Result Diversification. in *Proceedings of International World Wide Web Conference (WWW 2010)*, pp 881–890, 2010.

[81] Sato, S., Matsuyoshi, S., and Kondoh, Y.: Automatic Assessment of Japanese Text Readability Based on a Textbook Corpus, in *Proceedings of International Conference on Language Resources and Evaluation (LREC 2008)*, pp. 28–30, 2008.

[82] 佐藤朝美: 子どもの不思議を観察することから生まれるインタフェースデザインの提案, 日本子ども学会 チャイルド・サイエンス懸賞エッセイ, 2005.

[83] Schneider, K.: Children and Information Visualization Technologies, Interactions, Vol. 3, No. 5, pp. 68–73, 1996.

[84] Van der Sluis, F., Duarte Torres, S., Hiemstra, D., and Van Dijk, E.M.A.G., Visual Exploration of Health Information for Children, In *Proceedings of European Conference on Information Retrieval (ECIR 2011)*, pp. 788–792, 2011.

[85] 角 薫, 長田瑞恵, 田中克己: アニメーションメディア変換システム Interactive e-Hon における親子エージェント情報提示モデル, 日本知能情報ファジイ学会誌, Vol. 18, No. 2, pp. 240–250, 2006.

[86] Sumi, K. and Nagata, M.: Interactive e-Hon as Parent-Child Communication Tool, *Human Interface and the Management of Information*, Vol. 6772, pp 199–206, 2011.

[87] 田村隆宏: 幼児と大人のカテゴリー化における特徴情報の役割, 教育心理学研究, Vol. 42, No. 3, pp. 306–314, 1994.

[88] Thompson, K.C., Bennett, P.N., White, R.W., Chica, S., and Sontag, D.: Personalizing Web Search Results by Reading Level, in *Proceedings of ACM International Conference on Information and Knowledge Management (CIKM 2011)*, pp. 403–412, 2011.

[89] 湯浅千映子: 子ども向け文章の情報の配列—「小学生新聞」を対象に—, 文体論研究, Vol. 52, No. 52, pp. 41–56, 2006.

[90] 湯沢正通: 幼児による階層的カテゴリーの状況的理解, 日本教育心理学会総会発表論文集, Vol. 39, No. 4, pp. 373–38, 1991.