



ピタリエ：興味・発達段階にピッタリの絵本を見つけます

NTTコミュニケーション科学基礎研究所

はっとり たかし こばやし てっせい ふじた さなえ おくむら ゆうこ あおやま かずお
服部 正嗣 / 小林 哲生 / 藤田 早苗 / 奥村 優子 / 青山 一生

NTTコミュニケーション科学基礎研究所では、子どもの興味や発達段階にピッタリの絵本を見つけることのできる検索エンジン「ピタリエ」を開発しました。ピタリエは、検索アルゴリズム、自然言語処理という情報科学の研究成果に加えて、人間科学の一分野で子どもの語彙獲得メカニズムの解明を目的とする発達心理学の知見を取り入れることでユニークな機能を実現しています。ここでは、ピタリエの機能とともに、各要素技術について紹介します。

幼児の発育に絵本が果たす役割

絵本の読み聞かせは子どもの言語発達を促進します⁽¹⁾。また、絵本を読めば、ハッピーエンドのお話、悲しいお話、教訓話などの多様なストーリーを疑似体験できるため、読み聞かせは子どもの情操教育にも寄与していると考えられます。子どもの興味や発達段階に合った絵本を見つけてあげることができれば、言語発達・情操教育の両面ですらに高い効果が期待できます。

インターネットでの絵本探しの現状

インターネットの検索エンジンを使えば作者やタイトルから絵本を検索できます。また、絵本に特化した購買サイトも複数存在し、年齢や性別などを指定するとその条件でオススメの絵本を探したり、最近の絵本売上ランキングを調べたりすることもできます。これらの絵本購買サイトには、購入者による絵本の感想が多数寄せられており、これらの感想を参考に絵本を選ぶことも可能です。さらには、自分と似た購買履歴を持つユーザが購入し、自分はまだ購入していない絵本を教えてくれる協調フィルタリングという機能により絵本の推薦を受けることもできます。似た購買行動をす

る人は同じ趣味を持つ可能性が高いので、ランキング上位のものよりも趣味に合った絵本を見つけていることができるかもしれません。

どうやって絵本を選んでいる？

前述のようにインターネットを利用した絵本探しには多様な方法があるのですが、絵本ミュージアムという絵本の大規模イベント会場で私たちが行ったアンケートからは意外な事実が明らかになりました。絵本の探し方、選び方について770組の親子に質問をしたところ、インターネット上で利用できるさまざまな方法を使って絵本を選んでいる親子はわずか188組で全体の24%程度に過ぎなかったのです。もっとも多かった回答は「書店や図書館に直接出向いて選ぶ(540組, 70%)」でした。インターネットが一般的になり、書店や図書館に行くことなく情報が手に入るようになって、実際に見て選ぶという従来型の親子が大勢を占めていたのです。

ピタリエがめざす絵本探し

現状インターネットで利用することのできる絵本探し技術だけでは、絵本を選ぶ決定打にはなっていないことが分かりました。しかし一方で、書店や図書

館には展示スペースの物理的制約があるため、書店や図書館に出向くだけでは出会うことのできる絵本が限られてしまっています。ピタリエは、従来の絵本探し技術では困難だったいくつかの機能を実現することで、子ども1人ひとりの興味や発達段階に合う絵本を見つけられる絵本検索エンジンとなることをめざします。ピタリエの全体像を図1に示します。

コンピュータには絵本の文章が難しい

既存の絵本探しサイトでは、絵本の内容に関する検索を行うために、実際に人が絵本を読んでその内容について主要なカテゴリに絵本を分類するアプローチが取られてきました。この方法は、人手でカテゴリ分類を行うため確実ではありますが、手間がかかるということと事前に設定されたカテゴリ以上の情報を得ることが難しいという2つの課題がありました。

ピタリエでは、自然言語処理という、コンピュータに文章を扱わせるための技術を用いて、絵本の文章を機械的に詳しく分析する機能が実装されています。この実装において、従来の自然言語処理技術をそのまま絵本の文章に適用することは困難であったため、新しい技術を開発する必要がありました。面白いこ

とに、絵本の文章は、人にとっては大人向け文章より簡単なのですが、コンピュータにとっては大人向けの文章よりも難しいのです。どうということかという以下のとおりです。自然言語処理の第一歩として、形態素解析という、文章を形態素（単語）ごとに分割する処

理があります。大人向けの文章は図2のように、ひらがなと漢字が入り混じっているため、ひらがなと漢字の境目を形態素解析のヒントとして利用できるうえ、漢字によって形態素の候補が限られます。ところが、絵本の文章はほとんどひらがなしかないため、境目のヒント

が使えず、形態素の候補も非常に多くなるため、従来法による形態素解析が困難になります。

そこで私たちは、絵本の特徴に合わせて学習データと辞書を自動構築することで、ひらがなが主に出てくる文章に対して高精度な形態素解析を可能に

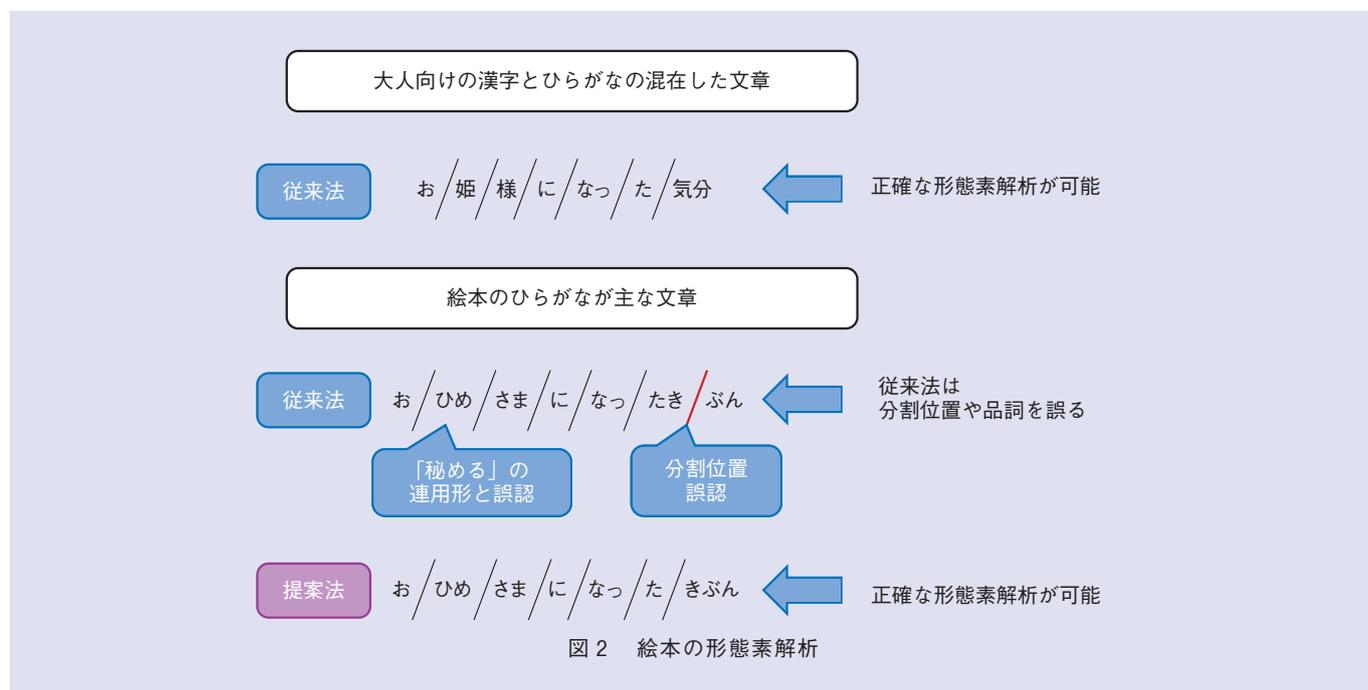
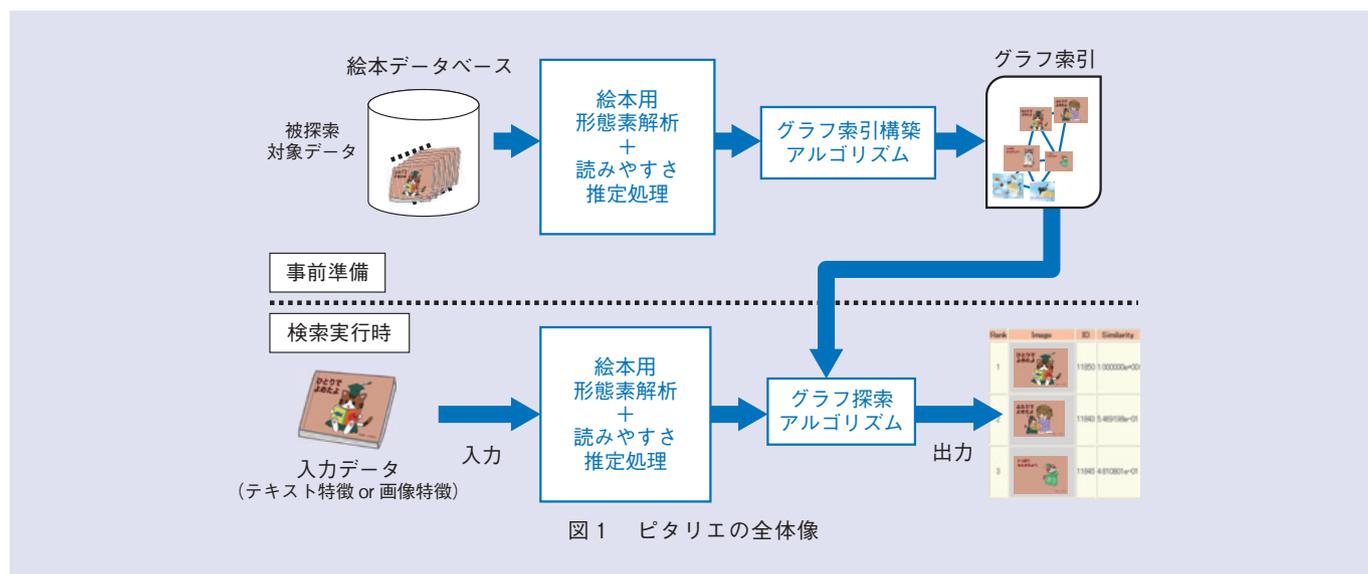




図3 ピタリエの検索結果画面

しました⁽²⁾。この技術を利用して、ピタリエは絵本一冊一冊にどんな言葉が使われているかを検索したり、対象年齢を推定したりできます。

興味のある内容の絵本を探す

私たちは絵本の特徴として、絵本一冊一冊に現れる言葉の種類とその頻度、および著者などの書誌情報という2種類のテキスト特徴に注目しました。絵本の内容に注目して検索する場合、これらが重要だと考えたからです。乗り物に関する絵本には、乗り物に関連した言葉が種類、回数ともに多く現れますし、著者によって書かれるお話の傾向が異なる場合もあります。

ピタリエでは、これらのテキスト特徴に基づいて類似探索という種類の検索を行います。従来の絵本検索システムは、入力した少数の情報に厳密に一致する絵本を見つけることが得意です。これに対して類似探索は、入力した大量の情報に基づいてできるだけ多くの条件を満たすという検索を得意としています。厳密な検索は、『ごんぎつね』というタイトルの絵本を探す』『新美南吉』が書いた絵本を探す』といった、最初から探したいものが明確な場合の絵本探しには向いています。しかし、探したいものが明確でない場合には類似探索のほうが適しています。従来の検索方法では難しい、「お気に入りの絵本に傾向が似た絵本を見つけたい」「タイトルが思い出せないけれど確か家族で海に遊びに行つて誰かが溺れそうになる絵本を見つけたい」などという検索を実現することができます。

『ぐりとぐらのかいすいよく（なかがわりえこ作・やまわきゆりこ絵、福音館書店、1976年）』と入力してピタリエで検索を行った検索結果画面を図3に示します。この場合、実際に内部で検索

表1 図3に引用した絵本の出典

タイトル	著者・挿絵画家など	出版社	出版年
ぐりとぐらのかいすいよく	なかがわりえこ（作）やまわきゆりこ（絵）	福音館書店	1976
ぐりとぐらとくるりくら	なかがわりえこ（作）やまわきゆりこ（絵）	福音館書店	1987
ぐりとぐら	なかがわりえこ（作）おおむらゆりこ（絵）	福音館書店	1963
ぐりとぐらとすみれちゃん	なかがわりえこ（作）やまわきゆりこ（絵）	福音館書店	2000
ぐりとぐらのえんそく	なかがわりえこ（作）やまわきゆりこ（絵）	福音館書店	1979
ぐりとぐらのおきゃくさま	なかがわりえこ（作）やまわきゆりこ（絵）	福音館書店	1966
ぐりとぐらのおおそうじ	なかがわりえこ（作）やまわきゆりこ（絵）	福音館書店	2002
こんにははまたおてがみです	中川李枝子（作）山脇百合子（絵）	福音館書店	2014
めがねうさぎのうみぼうずがでる	せなけいこ	ポプラ社	2005
なぞなぞえほん2のまき	中川李枝子（作）山脇百合子（絵）	福音館書店	1988
なぞなぞえほん1のまき	中川李枝子（作）山脇百合子（絵）	福音館書店	1988
なぞなぞえほん3のまき	中川李枝子（作）山脇百合子（絵）	福音館書店	1988
ぐりとぐらのあいうえお	なかがわりえこ（作）やまわきゆりこ（絵）	福音館書店	2002
ぐりとぐらのしりとりうた	なかがわりえこ（作）やまわきゆりこ（絵）	福音館書店	2009
ぐりとぐらのおまじない	なかがわりえこ（作）やまわきゆりこ（絵）	福音館書店	2009
うみ	中川ひろたか（文）はたこうしろう（絵）	自由国民社	2011
ありがとうともだち	内田麟太郎（作）降矢なな（絵）	偕成社	2003
10ぴきのかえる うみへいく	間所ひさこ（作）仲川道子（絵）	PHP研究所	2004
チリとチリリ うみのはなし	どいかや	アリス館	2004
ねずみのおいしゃさま	中川正文（作）山脇百合子（絵）	福音館書店	1977
そらいろのたね	なかがわりえこ（作）おおむらゆりこ（絵）	福音館書店	1964

されているのは、『ぐりとぐらのかいすいよく』に登場する195種類の言葉とその頻度、そして著者や出版社などの5種類の書誌情報という多数のテキスト特徴です。ピタリエに蓄えられた2012冊の絵本データベースの中から、これらの200種類のテキスト特徴について、できるだけ多くを共通して持つ絵本が検索結果として出力されています。同じ著者が書いたという点で似た絵本もあれば、「海」「泳ぐ」「浮き輪」などの言葉が共通する絵本もあります。

このように、入力している情報が大量なので、結果もいろいろな特徴に基づいて類似したものが出力されており、複雑な結果となっています。この複雑な検索結果を直感的に理解するため、「絵本の地図」機能があります(図4)。ピタリエでは、高速な検索を実現するための索引として類似の絵本どうしが結合したグラフ構造(ネットワーク構造)を用いて検索を行います。これをグラフ索引型類似探索法と呼びます。「絵本の地図」機能は、このグラフ索引のうち、検索結果に関連した部分を表示します⁽³⁾。同じ特徴に基づいて検索された絵本どうしは地図中で近隣に配置されるため、図4の例では検索結果が大まかに2通りのグループに分かれています。

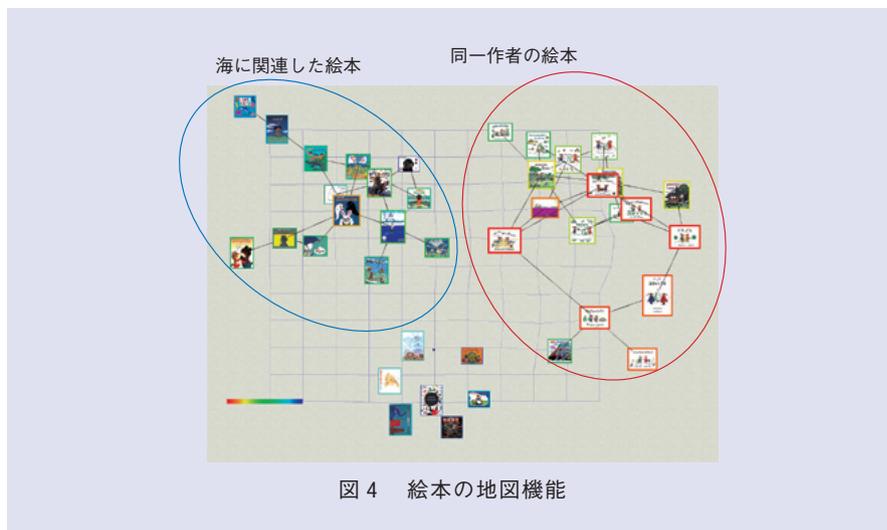


図4 絵本の地図機能

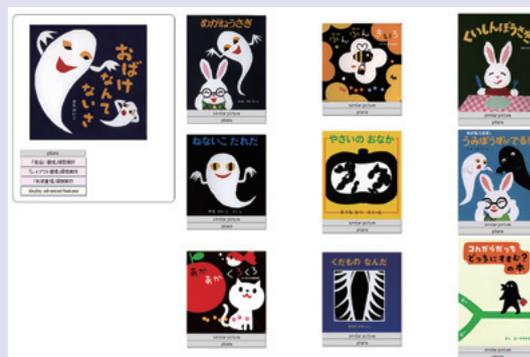


図5 表紙の絵で検索した結果画面

絵で探す

グラフ索引型類似探索法は、検索する対象どうしの中に何らかの「距離」*を定義することができれば適用可能な汎用的な探索法です。絵本の絵の類似関係に基づいて検索を行うことで、自分の好きな絵に傾向が似た絵本を探すことができます。ピタリエを用いて『おばけなんてないさ(せなけいこ作, ポプラ社,

表2 図5に引用した絵本の出典

タイトル	著者・挿絵画家など	出版社	出版年
おばけなんてないさ	せなけいこ	ポプラ社	2009
めがねうさぎ	せなけいこ	ポプラ社	1975
ぶんぶんきいろ	柏原晃夫	学研教育出版	2010
くいしんぼうさぎ	せなけいこ	ポプラ社	2004
ねないこだれだ	せなけいこ	福音館書店	1978
やさいのおなか	木内勝	福音館書店	1997
めがねうさぎのうみぼうずがでる!!	せなけいこ	ポプラ社	2005
あかあかくるろ	柏原晃夫	学研教育出版	2010
くだものなんだ	木内勝	福音館書店	2007
コンガラガっちどっちにすすむ?の本	ユーフラテス	小学館	2009

* 正確には、距離公理を満たさない非類似度でも可能。

2009年)』の表紙に似た絵の傾向がある絵本を検索した結果を図5に示します。

絵の傾向と一口に言ってもさまざまな種類が挙げられます。例えば、色遣いに注目すると、シンプルな色遣いでベタ塗りの絵本と、色が段階的に変化するグラデーション部分の多い絵本があります。色遣いの傾向を扱うために、私たちは絵に使用されている色の分布(カラーヒストグラム)という画像の特徴に注目しました。図6に示すように、ベタ塗りの絵のカラーヒストグラムは一部の色が集中的に使われるため所々が針のように切り立った形をしているのに対して、グラデーションが多い絵のカラーヒストグラムはなだらかな山の稜線のような形になります。また、絵にどのくらい

描き込みがなされているのかについても絵の特徴の1つとして採用しました。

エッジ検出という手法を用いると、図7に示すように絵の中から線を抽出することができます。描き込みが多くなされた複雑な絵は絵全体に占める線の割合が多く、デフォルメされたキャラクタが登場するシンプルな絵は線の占める割合が小さくなります。さらに、絵のどの領域に線が存在しているかの分布を見ることで全般的に描き込まれた絵なのか、それともキャラクタなどの一部が集中的に描き込まれた絵なのかを知ることができます。このように、絵の複数の特徴を複合的に利用することで、似ている絵の絵本を検索することができます。

発達段階に合った絵本を探す

出版社によって、絵本に対象年齢情報が付与されている場合があります。絵本のプロフェッショナルによって付与される、子どもの発達段階にあった絵本を探すのに有用な情報です。しかし、ベストセラー絵本を中心に2000冊以上の絵本について調べたところ、対象年齢情報が付与されていたのはそのうちの半分弱でした。さらに、付与されている対象年齢情報が「赤ちゃん向け」「1・2・3歳向け」のようにあいまい、あるいは幅をもたせた表現であることが少なくありません。子どもは2歳以降になると1カ月で覚える語彙の数が増え、「2歳になりたて」と「もうすぐ3歳」では同じ2歳という年齢であっても習得語彙数が大きく異なるため、対象年齢情報のみでは発達段階に応じた絵本選びは困難であると考えられます。

そこで私たちは、発達心理学の知見と自然言語処理の技術を合わせて絵本の文章の読みやすさを推定する技術を開発しました。発達心理学とは、人間の発達に関する学問で、私たちは特に、子どもの語彙獲得メカニズムの解明に注力しています。その一環として、延べ3000組以上の親子にご協力いただき、いつどんな言葉を子どもが発話できるかに関するデータを収集しています⁽⁴⁾。語彙獲得には少なからず個人差が出るのですが、得られた実験データからロジスティック回帰を行い、それぞれの言葉を半数の子どもが発話する時期を正確に求めることに成功しました。この語彙獲得時期情報と、出版社が付与した対象年齢情報と相関のある「一文の長さ」などの文章の特徴から、文章の読みやすさを推定できる技術を開発しました⁽⁵⁾。この技術を用いれば、出



図6 カラーヒストグラムによる検出

表3 図6に引用した絵本の出典

タイトル	著者・挿絵画家など	出版社	出版年
こぐまちゃんありがとう	わかやまけん	こぐま社	1972
じどうしゃ	わかやまけん	こぐま社	1994
いつまでも	アンナ・ピンヤタロ	主婦の友社	2007



図7 Canny法によるエッジ検出

版社が対象年齢を付与していなくても対象年齢を推定することもできますし、同じ対象年齢情報が付与されている絵本に対しても読みやすさによって順序を付けることもできます。また、お気に入りの絵本に近い読みやすさの絵本を見つけることができます。ピタリエを用いれば従来よりも細やかに発達段階に合った絵本を探していただけます。

今後の展開

ピタリエは最新の情報科学の技術に加え、子どもをよりよく理解するための人間科学の知見を取り入れて開発されています。結果として、従来の絵本検索システムでは困難だった「お気に入りの絵本に内容や絵が似ている絵本の検索」「あいまいなストーリーの記憶に基づく絵本の検索」「対象年齢情報が付与されていない絵本の文章の読みやすさの推定」などの先進的な機能を実現することができました。

今後は、図書館を舞台とした試用実験を予定しています。実際に多くの親子に使っていただくことで現状の課題とさらなる目標を洗い出し、さらにピッ

タリの絵本を見つけられる新バージョンの研究開発に取り組みます。また、ピタリエを構成するグラフ索引型類似探索、自然言語処理などの要素技術の深化についても取り組んでいきます。

参考文献

- (1) G. J. Whitehurst, F. L. Falco, C. J. Lonigan, J. E. Fischel, B. D. DeBaryshe, M. C. Valdez-Menchaca, and M. Caulfield: "Accelerating language development through picture book reading," *Developmental Psychology*, Vol.24, No.4, pp.552-559, 1988.
- (2) 藤田・平・小林・田中: "絵本のテキストを対象とした形態素解析," *自然言語処理*, Vol.21, No.3, pp.515-540, 2014.
- (3) 服部・青山: "グラフ索引を用いた絵本の類似探索～特徴の融合と結果のグラフ可視化～," *情報処理学会研究会 第10回 ネットワーク生態学シンポジウム*, 2013.
- (4) 小林・奥村・南: "語彙チェックリストアプリによる幼児語彙発達データ収集の試み," *信学技法*, HCS2015-59, pp.1-6, 2016.
- (5) 藤田・小林・南: "幼児を対象としたテキストの対象年齢推定方法," *認知科学*, Vol.22 No.4, pp.604-620, 2015.



(後列左から) 奥村 優子/ 小林 哲生/
藤田 早苗
(前列左から) 服部 正嗣/ 青山 一生

情報科学と人間科学の融合により、子どもの発達を緩やかに後押しする技術を提案するとともに、各要素基礎技術の深化に努めます。

◆問い合わせ先

NTTコミュニケーション科学基礎研究所
協創情報研究部 知能創発環境研究グループ
TEL 0774-93-5335
FAX 0774-93-5155
E-mail hattori.takashi@lab.ntt.co.jp