インターネット上のソフトウェア資源検索システムの設計と評価

広瀬雄二[†] 大駒誠一[††]

我々はインターネットから非常に多くの資源を得て利用している．ソフトウェアやデータなどのパッケージもそれらの資源のうちの1つである．現在，インターネット上に散在する無数のパッケージの中から欲しいものを探すためのサービスとしてarchieが存在する．しかし，archieは検索キーとしてパッケージのファイル名しか指定できないため，探しているもののファイル名を知らない場合や，探しているソフトウェアに関する知識がはっきりしていないような場合などには無力である．本論文では，ファイル名だけではなく，カテゴリ，OS，マシン，用途などパッケージの持っている属性を検索キーとして指定できるコンテンツベースの検索システムを提案する．また，それを運用することにより得られた結果について検討しシステムを評価する．

Design and Evaluation of the Retrieving System for Software Resources in the Internet

YUUJI HIROSE[†] and SEIICHI OKOMA[††]

Thanks to the growth of the Internet technology, we can obtain various software or data packages from the network. We currently use *archie* to search the enormous number of packages for the suitable one to us. However *archie* allows only pattern of packages' file name for the searching key. Therefore we cannot search any package with *archie* when we have no information about file name of packages or when we are looking for some package without the firm conscious of what we actually need. In this paper, we propose the contents-based package retriever which allows attributes and characteristics such as category, OS, machine, and their use, for searching keys.

1. はじめに

インターネットの普及により我々はおびただしい量の情報をいながらにして簡単に入手できるようになったばかりでなく，自分の持っている情報を不特定多数に対して供給することも容易になった．インターネット上での数あるサービスの中でもWWW（World Wide Web）は，その構築のコストの低さと視覚的効果の高さにより瞬く間に世界中に広まった．これにともないWWWにより提供される情報を検索するための技術も広く研究の対象にされ，実用的なシステムが数多く登場した．そのおかげで我々は，情報洪水の中で適切に知りたい情報を得ることができるようになった．

一方，インターネット上で流通している資源の1つ「ソフトウェア」の検索はどうかというと1992年から利用されるようになったarchieが現在でも主流サービスの位置を占めたままそれ以上のものは普及していない．

一般的に，ソフトウェア，データ，あるいはドキュメントなどのパッケージは，`tar+gzip`や`zip`などアーカイバと呼ばれるツールを用いて**アーカイブ**という一塊のファイルに納められanonymous ftp上で公開され，配布される．

archieは，そうしたアーカイブのファイル名リストを，集中データベースサーバとなるarchie serverに登録しておき，archie専用検索クライアント（archieコマンドなど）にファイル名をキーとして与えることにより，目的アーカイブの位置情報を解として与えるシステムである．

archie serverは，各anonymous ftpサービスホストのファイル名リストを定期的に回収し自己のデータベースに登録する．どのホストのリストを集めるかは各archie serverにより決まっているので，当然のことながら検索対象となるアーカイブはいずれかのarchie

† 慶應義塾大学理工学研究科管理工学専攻
Department of Administration and Engineering, Faculty of Science and Technology, Keio University
†† 慶應義塾大学理工学部管理工学科
Department of Administration and Engineering, Faculty of Science and Technology, Keio University

server に登録してある anonymous ftp サービスホストに存在するものに限られる．

本論では，パッケージ検索者を，検索するときに持っている動機と情報のレベルによって分類する4層モデルを導入し，これまでのサービスでは解を与えることのできなかった層が存在することに言及する．そのうえで，それらの層に対する解を与える機構を提案し，その有効性を示す．

2. パッケージの検索

我々がネットワーク上から何かを得ようとする場合，求めたい資源が何らかの文書に記述された「情報」である場合には，検索対象文書に探したいことがらに関する言葉が含まれていることが明らかなので，

探すためのキーワード ∈ 知りたいこと

という概念上の関連が約束できる．ところが，アーカイブ検索においてはこのことはまったく保証されない．この理由としては以下があげられる．

- 検索のためのキーがファイル名に限られている
- ファイル名がアーカイブの中身の性質を表しているとは限らない
- ファイル名長等（8+3 文字等）の制限があり，意味を持たせることが困難である

これらの制約は，検索者がどのようなキーワードを選択すればよいかを決定するのを非常に困難なものとしている．にもかかわらずいまだに archie が有効に利用されているのは，NetNews の力に負うところが大きい．これについては 2.1.1 項で説明する．

以上のように，通常テキストの検索と，アーカイブ検索では検索者の発すべき問合せを導出する過程に差異があるため，両者を同系列で扱うことはできない．ここでは，パッケージ検索者がなんらかのパッケージを必要とする状況に至る経緯を4つの層に分類し，各層に対して異なる解発見アプローチをとるべき必要があることについて述べる．

2.1 検索動機の四層分類

アーカイブ検索では，検索者本人が検索対象物に関してどの程度の知識を持っているかによって検索操作に必要な性質が大きく異なってくる．検索者の目的アーカイブに関する知識の度合に応じてその対象を求める動機も変化し，それは以下の4段階に分類できる．

(i) ある問題を解決したい．
(ii) ある機能を提供するパッケージが欲しい．
(iii) 名前の分からない特定のパッケージが欲しい．
(iv) 名前の分かる特定のパッケージが欲しい．

レベル i は，検索者が現在計算機上で対象としている問題を解決するうえで，漠然と不便であるとか，何らかのトラブルがあるだとかの問題をかかえてはいるもののその具体的解決策は知らない，または思い付かない場合に生ずる動機である．

通常このレベルの動機を持っている検索者は，手がかりが得られないので，一般的な検索システムに対する問合せを導くことができない．そのため自ずと曖昧な質問をするしかなく，回答を人間に求めて，問合せを身近にいる詳しい人，あるいは NetNews などに発する．

レベル ii は，パッケージ自体に求める機能ははっきりしているが，そのような機能を有するパッケージが存在するかどうかは定かでない状態である．たとえば，あるユーザが S というソフトウェアを O_1 という OS 上で有効に利用している場合に，別の OS で S の移植版，もしくは同等の機能を持ったソフトウェアを利用したいと思ったが，実際にそれが存在するかどうかは分からない，といった場合などがこれにあたる．

ただし，どのような機能を必要とするか自体の発想が誤っている可能性もあり，たとえ検索が成功して何らかの解が得られたとしてもそれが本当に当該問題を解決するために有効なパッケージを指しているどうかかは保証されない．

レベル iii は，パッケージ自体に求める機能，およびそれを有するパッケージが存在することが分かっていて，その名前のみを失念している状態である．

レベル iv は上の状態に加え，パッケージの名前，またはその一部があらかじめ分かっている場合で，すでにそのパッケージの古いバージョンなどを持っているような場合に相当する．

2.1.1 各レベルの検索に必要なもの

先述したとおり，各レベルの検索者が持っている知識，必要としている情報はそれぞれ異なる．それゆえ全レベルの検索行為を一律に扱うことはできない．ここに各レベルの検索者に供給すべきサービスが備えるべき性質を示す．

レベル i レベル i の検索者は，発生した問題を解決してくれるようなパッケージを求める．このような場合同様な問題を解決した人からの回答を集めた症例データベースが必要である．ただし，検索者自身が適切な問合せを発行することも期待できないので，データベースを構築したとしてもそれを有効に活用することができるかどうかは疑わしい．

レベル ii レベル ii の検索者は，パッケージの持っている機能からそのパッケージのアーカイブ名を

索くことができればよい．したがって，パッケージの持っている機能・性質を端的に表したファイル（以後紹介文書）を用意し，それらの中から検索をしてマッチするキーワードを含む紹介文書を持つものが求めるパッケージであるという解を返せばよい．

ただし，検索者が期待した解決方法を提供するパッケージが存在することもあるが，それとはまったく違ったアプローチで解決方法を提供するものが存在することがある．たとえばUNIXオペレーティングシステムを用いている場合に，誤ってファイルを消してしまった場合検索者は，「消去してしまったファイルを復活するツールはないのか」という動機から検索を試みる．ところが現在普及しているUNIX系OSではファイルを復活することは現実的には不可能で，通常は誤って消去した場合に備えてこまめにバックアップをとったり，ファイルを消去するコマンドの代わりに一時的なゴミ箱ディレクトリに移動するツールを用いることで不慮の誤消去に対処する．

したがって，UNIXを利用している人が，ファイルを復活するためのツールが欲しいという問合せを発した場合は，バックアップを容易にとるためのツールを解として返す，といった変換が必要である．一般的には，解決不能な問合せを解決可能なものに変換したうえで検索動作を起こすような機構が必要となると言い換えられる．

レベル iii このレベルの検索者は単にパッケージの名前が分からないだけで，求める機能もはっきりと表現できるレベルにある．したがって，紹介文書を用意すればそれらの中から適切なキーワードで検索させることが可能である．

レベル iv すでにパッケージの名前が分かっているので既存サービスarchieを使えばよい．

レベルi～iiiに関しては，現状では問合せを受け付ける適切なサービスが存在しない．このため検索者はNetNewsに質問を投稿し，解を知っている別の読者がレベルivの問題にまで変換したうえで，回答を提示するという過程をとっている．archieが現状でも頻繁に使われている理由はまさにここにあり，NetNews上のやりとりがarchieの適切なフロントエンドとして有効に機能しているからである．しかし，ときには逆にNetNewsに質問を投げかけること自体，検索者にとって敷居が高い場合もあり，検索を断念してしまうこともある．

3. Dartsの設計

すべてのレベルの検索者の問合せを名前ベースの検索サービスであるarchieに頼っている現状に対し，コンテンツベースの検索を可能にするシステムDartsを提案する．

3.1 基本思想

アーカイブ検索システムを構築するうえでの基本理念として，

(i) 4レベルいずれに属する検索者に対しても解を与えられること

(ii) データ構築のコストがかからないこと

ということを掲げた．

archieのカバーできない検索者層をカバーするのが目的であるゆえiは必須である．またiiに関して，archieがこれまで利用され続けてきた理由の1つに，データの自動構築性の高さがあげられる．archieサーバに置かれるアーカイブのリストはすべて定期的に自動取得されるものであるので，検索者側からはデータの新規性がつねに保たれることになる．また管理者側から見た場合も，運用時の労力が皆無に等しいのでサービス提供を行うことが容易である．これらのことが相乗効果をもたらしarchieを標準的なサービスたらしめた．したがってDartsにおいても検索用のデータベース構築を自動化できるよう，インターネット利用者のactivityを積極的にDartsのデータに取り込む方針で設計した．具体的には，

- NetNewsで交わされる質問・回答記事の利用
- Darts利用者のインタラクションからの紹介文書の抽出

という手法を取り込むことにより，管理者のデータ構築コストを低減させることを目指している．

Terveen[6]らは，NetNews記事中のうちhttp://を含むものを選び，さらに字面解析によりURLの紹介である記事を抽出，分類配置し，有用なWWWページを自動的に選出することに成功している．PHOAKSと呼ばれる彼らのシステムでは，WWW検索エンジンを利用して情報検索した場合の「適合するものが極端に多く表示される」という問題点を解決し，さらにはNetNewsにおける質疑応答記事が人間の持っている資源情報の宝庫であり，機械処理により抽出利用することが可能であることを実証している．ただし，PHOAKSは，WWWページの紹介文のみを収集対象としているため，パッケージを探す場合には，それ自身に関する情報を載せたWWWページが存在している必要がある．

一方，多くのパッケージを，その説明書をともに検索可能な形にしてサービス提供しているものがある．SHAREWARE.COM☆や PACK PROJECT☆☆などがその例である．これらはあらかじめ要求の高いと思われるパッケージを大量にデータベース化して所蔵しておく形式であるため，データベース構築に多大な外部記憶等を必要とするだけでなく，検索者が目的パッケージの持っている機能をはっきり把握している場合しか検索できないという問題があり，最も困っている，前提知識の低い検索者に対してのサービスが与えられないという問題点がある．

以上のことをふまえて Darts では NetNews の質疑応答記事自体を検索用データとして用い，紹介WWWページを持たないパッケージをも検索可能とするために ftp://に着目することとする．また，あらかじめ可能性のあるパッケージのすべての紹介文書を自動抽出するのではなく，検索者自らが読んだ紹介文書を Darts 自身のデータベースに取り込む形式を採用する．

なお，本論では NetNews 上で目的アーカイブの所在に関しての質問と回答を繰り広げている記事のことを Q/A 記事と呼ぶことにする．

3.2 Q/A 記事の特性

NetNews 上にはあらゆるレベルの検索者からの質問とそれに対する回答が流通している．これを検索用データベースとして利用することで同様の問題を抱えている人の問題解決の補助とすることができる．さらに，以下に述べる NetNews の特性により，全 NetNews 記事の中からアーカイブ検索のために有効な Q/A 記事を自動抽出することが容易になっている．

一般に，NetNews に対して発せられる質問記事と，回答記事は**図 1** のような木構造の参照関係が存在する．図における最初の質問に直接関わる回答記事は網かけで示した部分である．慣習的に NetNews 上でのアーカイブに関する質問に対する回答記事は，

(i) 元記事の質問文の引用

(ii) パッケージの所在を表す記述（URL☆☆☆など）

が含まれることがほとんどである．言い換えるならば，回答記事のみに質問と回答の要素両方が含まれているということで，これはすなわち，全 NetNews 記事の中から，ii が含まれる回答記事のみを抽出することで，検索に必要な情報ほぼすべてをカバーできることになるということである．また，ロケーションパターンは

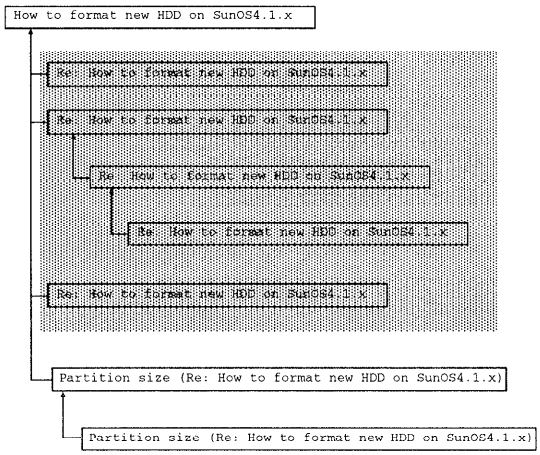


図 1 NetNews に投稿された記事のスレッド
Fig. 1 Thread of NetNews articles.

正規表現で記述できるのでそれ自身を含むファイルの抽出も容易である．逆に，質問記事に含まれることが期待できるのは自然言語による問合せ文のみであるため，膨大な量の NetNews 記事から質問記事を正確に自動抽出することは困難である．

以上の理由により Darts で利用する検索用データとしては，NetNews 中の全 Q/A 記事のうち，参照木の根（すなわち質問記事そのもの）は無視することにし，ロケーションパターンを持つ回答記事のみを抽出して構築するデータベースに採り入れることとする．

3.3 システム構成

Darts では 4 つのレベルすべての検索に対する解を与えられることを目的としているが，レベル iv の検索者に対しては単に既存のサービスである archie を呼ぶためのフロントエンドとして機能させる．レベル ii〜iii に属する検索は Darts 固有の検索部が解を与える．**図 2** が Darts 固有部の概念図である．

さらに以下が，Darts を構成するすべてのモジュールである．

位置情報抽出器 NetNews 記事からパッケージのロケーション情報を抽出して Q/A 記事スプール構築

Archie インタフェース レベル iv の検索者へのサービス

検索ボード Q/A 記事，紹介文書スプールからの検索

インタラクショントレーサ ユーザのアーカイブ閲覧時の動作から各種情報を取得しデータベース再構築のためのフィードバックを行う

紹介文書抽出器 インタラクショントレーサ

☆ http://www.shareware.com/
☆☆ http://www.vector.co.jp/
☆☆☆ 本論で取り扱う URL は種別 FTP（ftp://で始まるもの）のみに限定する．

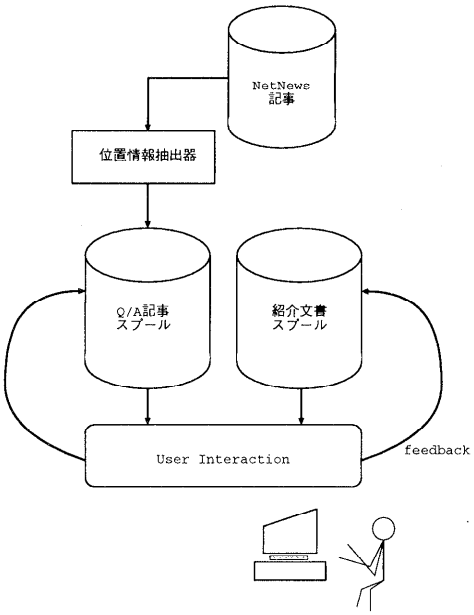


図2 Darts 概念図
Fig. 2 Structure of Darts.

の取得した紹介文書を既存データベースに追加

報告器 無効データファイルを管理者に報告

再構成器 Q/A 記事スプールの分類再構成

図2中の「NetNews 記事」のうち本システムで扱うものは fj ニューズグループすべての記事である．これらの記事のうちパッケージの在処の質問と回答を含む記事すべてを位置情報抽出器により抽出し，さらにそれぞれの記事を，その記事が言及しているパッケージの用途あるいはパッケージの動作する環境に応じて分類したうえで Q/A 記事スプールに格納する．

このスプールは，おもにレベル i～iii の問合せをカバーすることを想定している．また，Darts 利用者から分類が不適切と思われる記事があるという報告を受け取った場合には再構成器により自己再編成が行われる．

紹介文書スプール は，インタラクショントレーサと紹介文書抽出器が，Darts 利用者の検索時の行動から各種パッケージの紹介文書を集めたもので，おもにレベル ii～iii の問合せをカバーすることを想定している．ユーザが与えたキーワードをそれぞれのデータから検索し，マッチする単語を含む文書をユーザに提示する．提示文書には該当パッケージへのリンクが張られており，ユーザはそのリンクをたどることで目的パッケージに到達できる．

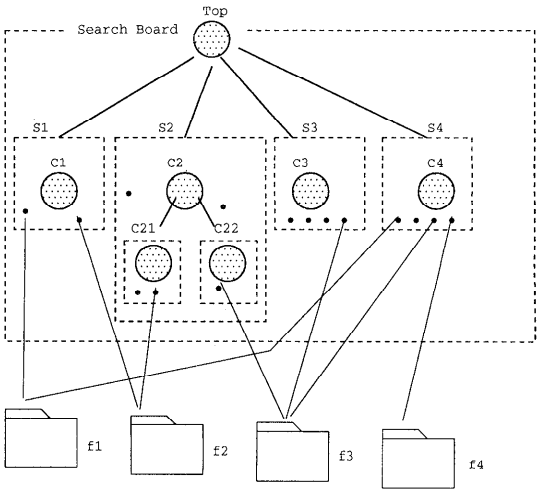


図3 分類木
Fig. 3 Categorization tree.

3.3.1 Q/A 記事スプール

NetNews から抽出された Q/A 記事群は図3のように，あるカテゴリ固有の概念を表す単語を持つ文書が，同一グループに属するように分類配置される．各文書は複数のグループに属することもあり，そのような場合であってもファイルシステム上では同一ファイルとして存在する．各グループには，グループ内の記事のみに範囲を限定した検索ボードが付属し，検索者があらかじめ検索対象を単一グループ内に制限することができるようになっている．たとえば，図中で文書 f1 はカテゴリ C1 と C4 の両方の概念を包含するものであるから両方のグループに属し，どちらの検索ボード（S1, S4）からも検索できるようになっているが，ファイルシステム上での実体は1つである．

この分類木を生成する位置情報抽出器は，分類表中の正規表現群と記事中に含まれる単語を照合し，いずれかの正規表現に適合するパターンがあった場合にその記事を対応するカテゴリに分類する．この分類表は図4のように，左辺にカテゴリ名，右辺にそのカテゴリ固有の概念を表す典型的な単語群を正規表現で記述したものを列挙したものである．

位置情報抽出器は図5のような2種3段階の分類を行っている．分類の第1および第2段階ではそれぞれ，記事に含まれる Newsgroup ヘッダと，Subject ヘッダのみを照合の対象として分類を行う．一般的に2つのヘッダには記事の内容にふさわしい単語が含まれていることが期待でき，仮にふさわしくない Newsgroup に投稿されていたり，ふさわしくない Subject が付けられていた場合には，回答者がふさわしいものに変え

```
archiver    \barchiver|(un|de)?compress|\b(zip|arc|arj|lha\|lzh|tgz
            |tar(\.(gz|Z))?)\b|解凍|binhex
faq         (FAQ|Frequently\s?Asked\s?Question)
textutil    convert|translate|code\s+conv|henkan|uu(de|en)code|\b(a2ps
            |psutil)
device      \b(hdd|mo|pd|disk|scsi|ide|pci|mouse|memory)\b|(tape|
            cd[ -]?rom|modem|fax|cd ?play|printer)
```

図4 分類表（抜粋）
Fig. 4 Categorization table.

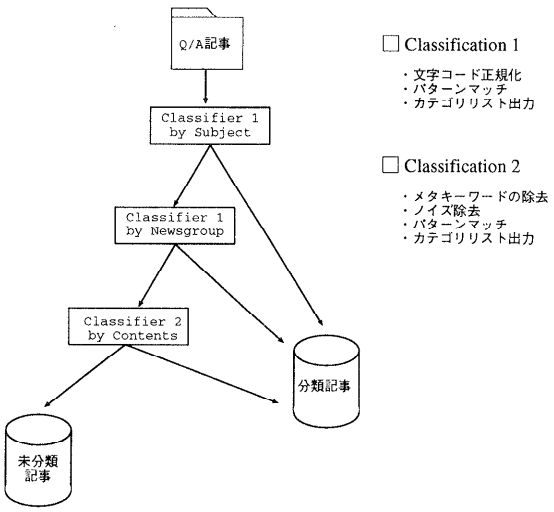


図5 分類の流れ
Fig. 5 Categorization.

てから回答を投稿することが多く見受けられる．

上記2つの段階の過程でどのカテゴリにも分類されなかったものを第3段階の分類にかける．ここでは照合の範囲を記事本文にまで広げて分類を行うが，前2段階とは異なり分類の際にメタキーワードとノイズの除去を行う．メタキーワードとは，たとえば「詳細はメイルで質問して下さい」の「メイル」のような伝達手段を表すキーワードで，言及しているパッケージの属性には関係なく登場する性質のものであるため，メタキーワードになりうる単語はすべて除外してから照合を行っている．ノイズとは，signature部に含まれたロケーションパターンや，たとえば「# Wnnのちょうしがわるいのでひらがなでしつれいします」などのような，文脈とはまったく関係のない文章のことである．これらも同様に照合時に除外する必要があるが，文脈を完全に把握することは非常に困難であるため，Dartsではsignature部の除去のみ行っている．

3.3.2 紹介文書

これは各パッケージの紹介文書を収集する部分であるが，インターネット上に存在するすべてのパッケージの紹介文書を用意することは事実上不可能といってよい．そこでDartsでは，Darts利用者のインタラクションから紹介文書スプールを構築する手法をとっている．

Dartsが提示したすべての解から検索者がパッケージを取得する際に，そのURLが指し示すものが本当に所望のものなのか定かでない場合，検索者にアーカイブに含まれるファイル群をWWWブラウザ上で閲覧させる．検索者はファイル群のうちパッケージの内容の説明があるものを探してそれを読んだ結果，要不要の決断を下す．「必要」との判断をもたらした文書はそのアーカイブの紹介文書，もしくはそれに準ずるものであると判断できるので，その文書をDartsの紹介文書スプールに格納し，将来の検索に備える．

4. 実　　装

Dartsは3.3節で述べた各モジュールを実装するPerl言語によるスクリプト群で構成される．各モジュールのうち，Q/A文書スプールを初期生成する位置情報抽出器はオペレーティングシステムのユーザプロセスとしてコマンドインタプリタ上から起動される．残りの部分はすべて利用者がWWWサービスデーモン☆にアクセスすることをトリガとして起動される．Darts本体をPerl言語で実装したことにより，DartsはPerlインタプリタとWWWサービスデーモンの動作するプラットフォームすべてで動作させることができる．

Darts固有の検索部，すなわちレベルi~iiiの検索者に対するインタフェースは**図6**のような窓口で始まり，検索したいパッケージの性質に応じていずれかのカテゴリを選択すると，そのカテゴリ内に視野を絞ることができる．検索者は一覧表示されたQ/A記事のサブジェクトを閲覧，または任意のキーワードでQ/A記事中を検索して表示させることにより，目的アーカイブに関しての質疑応答のある記事にたどり着き，記事中に含まれるURLをクリックすることにより目的アーカイブを入手することができる．

また，仮にQ/A記事中のURLが古かったり，誤

☆ 本システムではNCSA httpdを用いた．

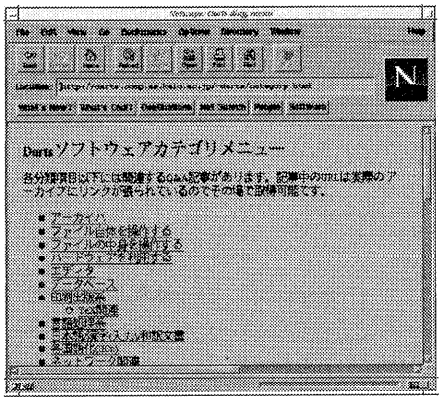


図6 Dartsカテゴリメニュー
Fig. 6 Category menu of Darts.

りがあった場合は，Darts自身が同名もしくはより新しいバージョンのアーカイブをarchieサービスを利用して検索し，その結果に置き換えたものを検索者に提示することで自己データベースの風化に対処する．

5. 検討と評価

5.1 Q/A記事分類木生成

1995年12月から1996年8月までの全fj NewsGroupのロケーションパターンを含む全記事を対象に分類を試みた．全Q/A記事が図5のどのフェーズで分類されたかの結果を以下に示す．

表1 のSubjectとNewsGroupによる分類率を見て分かるとおり，ヘッダ部分の照合だけで全体の87%のQ/A記事の分類が行われている．通常ヘッダにはノイズやメタキーワードが含まれにくいことを考えると，Q/A記事のかなりの部分が適切なカテゴリに分類されるということがいえる．実際にこれを検証するために分類された記事を無作為に100件選び分類カテゴリの正当性を確認した結果を**表2** に示す．表中 α，β，γ は以下のとおりである．

α = 正当なカテゴリのみへの分類
β = 正当なカテゴリのみへの分類と不当なカテゴリ両方への分類
γ = 不等なカテゴリのみへの分類

表2で明らかなように，SubjectとNewsgroupによる分類ではほとんどの記事が正当なカテゴリに分類される．表1の結果とあわせて概算すると，$0.585 \times 0.83 + 0.287 \times 0.87 + 0.068 \times 0.38 = 0.761$ となり，全記事のうち約76%が正当なカテゴリに分類されることが分かる．

仮にSubjectとNewsgroupにより分類された記事だけを採用した場合，全記事数4112(= 2759 + 1353)となり，$(2759 \times 0.83 + 1353 \times 0.87)/4112 = 0.843$ と，全記事の実に80%以上が正当にカテゴリ分けできるといえ，より高精度に分類されたQ/A記事スプールを構築できることも注目に値する．

表1 分類されたフェーズの割合
Table 1 Ratio of categorized phase.

全記事数	4715	—
Subjectによる分類	2759	58.5%
NewsGroupによる分類	1353	28.7%
記事内容による分類	391	6.8%
分類不能	212	4.5%

表2 分類成績
Table 2 Result of categrization.

	α	β	γ
Subjectによる分類	83	15	2
NewsGroupによる分類	87	7	6
記事内容による分類	38	25	37

表3 アクセスデータ
Table 3 Access data.

全アクセス	トップページ		3220
	全ページ		39217
archie検索	検索成功		1753
(Level iv) 2309	検索失敗		556
Darts固有検索	検索成功		1469
(Level i〜iv) 2258	検索失敗		789
目的アーカイブに到達 14976	Darts固有部から 10576	内容確認あり	5714
		内容確認なし	4862
	archieから 4400	内容確認あり	2721
		内容確認なし	1679

5.2 アクセス状況

Dartsを一般アクセス可能にした1995年末より現在までのアクセスデータについて分析する．

表3 によれば，名前ベースのarchie経由の検索とDarts固有のコンテンツベース検索がほぼ同じ頻度で動作していることが分かる．このことより，レベルi〜iiiに属する検索者の潜在的な多さが明らかになった．また，Darts固有検索時のキーワードを，その単語の持つ意味の性質ごとに分類した結果，パッケージの名前に属するものが全体の46%，それ以外のコンピュータ用語に属するものが46%，コンピュータとは関係ない一般用語が残りの8%であった．過半数がパッケージ名以外での検索を必要としていることからもコンテンツベースの検索が有効利用されていることが確認できた．

さらに，目的アーカイブ到達の項を見ると，archie経由でアーカイブを取得した場合の内容確認率が62%であるのに対し，Darts固有のサービス経由でアーカイ

ブを取得した場合の内容確認率が54%であることから，わずかではあるがコンテンツベースでの検索が，より目的アーカイブへの到達ステップを削減していることが伺える．

6. まとめ

本論文では，あらたに提案・実装したコンテンツベースのアーカイブ検索システムは，高精度に分類されたデータベースを自動的に構築することができるため，低いコストで運用できること，なおかつこれまでアーカイブ検索できるサービスを得られなかったレベルi～iiiの検索者に対して解を与えるサービスとして位置付けられることを示した．

現在Dartsは

`http://darts.comp.ae.keio.ac.jp/`

にて一般公開し，サービス供給している．1997年1月～5月の1日あたりの値は，約430件の有効HTTPアクセス，約35件の提示解へのアクセス，約20件の提示解の内容確認のためのアクセスという状況である．

現在実装済みの部分では，Darts自身で扱うデータを拡充させる方向の再構築しか機能しない．最新のフリーソフトウェアが数カ月のうちに陳腐化する可能性を持ったインターネットの世界ではURL情報の風化も早く，Dartsでも風化した情報を検出削除する機構が必要である．

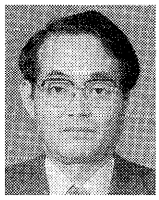
参考文献

1) Browne, S., Dongarra, J., Green, S. and Moore, K.: Location-Independent Naming for Virtual Distributed Software Repositories, *ACM-SIGSOFT Symposium on Software Reusability* (SSR'95).
2) Bowman, C.M., Danzig, P., Manber, U. and Schwartz, M.: Scalable Internet Resource Discovery: Research Problems and Approaches, *Comm. ACM*, Vol.37, No.8, pp.98–107, 114 (1994).
3) Berners-Lee, T.: Universal Resource Identifiers in WWW, RFC1630, CERN (June 1994).
4) Sollins, K.: Functional Requirements for Uniform Resource Names, *RFC1737*, MIT/LCS, (Dec. 1994).
5) Deutsc, P. and Emtage, A.: Publishing Information on the Internet with Anonymous FTP, *IETF Internet Draft* (May 1994).
6) Hill, W. and Terveen, L.: Using frequency-of-mention in public conversations for social filtering, HCI97.
7) Terveen, L.G., Hill, W.C., Amento, B., McDonald, D. and Creter, J.: Building Task-Specific Interfaces to High Volume Conversational Data, *SIGCHI CHI97*, Electronic Publications.

(平成9年5月12日受付)
(平成10年2月2日採録)

広瀬 雄二（学生会員）

昭和43年生．平成4年慶應義塾大学大学大学院理工学研究科管理工学専攻修士課程修了．現在同後期博士課程在籍．インターネット資源の処理，移動エージェントなどの研究に従事．



大駒 誠一（正会員）

昭和11年生．1959年慶応義塾大学工学部計測工学科卒業．同年小野田セメント株式会社入社．1964年慶応義塾大学工学部管理工学科助手．現在同理工学部管理工学科教授．工学博士．プログラミング言語，文字認識，コンパイラ，アルゴリズム，機械翻訳などに関する研究に従事．著書に「アセンブリプログラミング入門」（培風館），「改訂FORTRAN77」（サイエンス社），「数値計算のためのCプログラミング技法」（HBJ出版局），「COBOL基礎と応用」（サイエンス社）など．日本ソフトウエア科学会，計量国語学会，ACM，アジア太平洋機械翻訳協会各会員．