

4L-5

# インターネットにおけるソフトウェアアーカイブ位置情報の感応型提供\*

広瀬 雄二<sup>†</sup> 大駒 誠一<sup>‡</sup>  
慶應義塾大学理工学研究科管理工学専攻<sup>§</sup>

1997年9月25日

## 1 背景

大量の情報がWWW(World Wide Web)によって供給されている。これらのもののうち、文章主体の情報は各種のサーチエンジンを用いることで検索することが出来る。ところがソフトウェアのアーカイブは、それ自体がバイナリデータであるために対象物そのものを検索データとして用いることが出来ない。

本稿では、インターネット上において一般の文書を検索する場合とソフトウェアアーカイブを検索する場合についてそれらの性質の違いを考察し、効率的なソフトウェア検索サービス *Darts* の構築とそれを運用した上での検討点について述べる。

## 2 ソフトウェア検索者の四層モデル

一般の、文書/文献検索では、探したい文書に含まれそうな単語を指定すればよいのだが、ソフトウェアの場合は *archie* サービスを利用してファイル名をキーとして検索するのが一般的である。ところが、通常ファイル名は

- そもそもどういう名前がついているのか知らないケースが多い
- 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(DOSでは8+3文字)の制限があり、暗号のような名前が<sup>§</sup>ついていることが多い

\*Reactive Information Retrieval of Software Archive Location in the Internet

<sup>†</sup>Yuuji HIROSE

<sup>‡</sup>Seiichi OKOMA

<sup>§</sup>Department of Administration Engineering, Faculty of Science and Technology, Keio University

のような制約から、検索者がどのような検索キーを指定すれば良いか自体が非常に困難となってしまう。

ここで、パッケージ検索者が必要とするソフトウェアに対して持っている情報に応じて四つの層に分類するモデルを規定し、各層に対して異なる解発見アプローチを取るべき必要があることについて述べる。

### 2.1 検索者の四層分類

特定のパッケージがないかどうか探す時に検索者が、パッケージを要求する動機のレベルを分類すると以下ようになる。

#### I. 問題解決レベル

現在計算機上で対象としている問題を解決できるソフトウェアはないかと漠然と意識している状態。現状ではこのレベルの動機の検索者は、問い合わせを身近にいる詳しい人、あるいはNetNewsなどに発する。

#### II. 機能要求レベル

対象問題を解決するためのソフトウェアがどのような機能のものかははっきりしているが、それが存在するかどうかは定かでない状態。あるOSで既に使用しているソフトウェアと同等のものを別のOSで探す場合などがこれにあたる。

#### III. 名称未知レベル

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る、その名前のみを失念している状態。

#### IV. 名称既知レベル

レベル III 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、既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。

## 2.2 各層の検索に必要なもの

前述した四層の検索者が必要としている情報はそれぞれ異なる性質のものなので一元的なサービスで検索リクエストを賄うには無理がある。各レベルの検索者が必要としている情報をまとめると以下のようになる。

レベル I 発生した問題をどのようなソフトウェアで解決出来るのかという情報が必要としている。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。

レベル II 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が分かれば良い。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 short description)を用意し、それらの中からキーワード検索をさせれば良い。

レベル III このレベルの検索者は求める機能もはっきりと表現できることが期待できる。したがって、これも short description を用意すればよい。

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 を使えば良い。

## 3 Darts と検索者の層分類

### 3.1 ソフトウェアアーカイブ検索システム Darts

我々は上記四層全ての検索者の問い合わせに適切な解を返すソフトウェア検索サービス Darts を設計実装した。Darts は、NetNews 上で展開されるソフトウェアに関する Q&A 記事と、検索者が Darts を用いて検索し内容を調べたソフトウェアの short description をデータベースとして用いるシステムで、1996 年から稼働している。1997 年 6 月現在、一日平均 18 件のソフトウェアアーカイブの位置情報を解として返している。

### 3.2 検索キーワードの分析

これまで Darts では、レベル I~レベル III の検索者に対する解を一括して扱って来た。本稿では、Darts に検索キーとして与えられた単語を分析し、検索者の属する層の推定を試みた。

ソフトウェアを検索時に入力したキーワード全 1197 単語を調査したところ、それらは大きく分けて以下の 4 つに分類できることが分かった。

	個数	比率
1. ソフトウェア名に関するもの	554	46%
2. コンピュータ用語	420	35%
3. コンピュータ/OS の種別	129	11%
4. コンピュータとは無関係	94	8%

4 はソフトウェア以外(データなど)を探そうとしていたものなので除外するとして、1~3 は検索者の持っている知識が反映されていると考えることができ、これをシステム側の検索時情報として用いることが可能である。

## 4 まとめ

本稿ではソフトウェア検索時の検索キーワードをもとに、検索者の属する層を判別し、それに応じて検索データベースを切り替えることによって適切な検索対象空間を切れることが分かった。今後は、検索時の操作履歴を加味することなどにより、より正確な層分類を可能にしていきたい。

## 参考文献

- [1] Shirley Browne, Jack Dongarra, Stan Green, Keith Moore; Location-Independent Naming for Virtual Distributed Software Repositories; ACM-SIGSOFT Symposium on Software Reusability (SSR'95)
- [2] C. Mic Bowman, Peter Danzig, Udi Manber, and Michael Schwartz; Scalable Internet Resource Discovery: Research Problems and Approaches; Comm. of the ACM, Vol. 37, no. 8, pp. 98-107, 114(Aug 1994).