

26 WSRP リモートポートレット検索エンジン

伊藤真澄* 奥野拓 大谷真

(北大情報科学)[†]

1 はじめに

現在、インターネット上には、ニュースや株価、天気予報などの各種コンテンツ（ポートレット）へアクセスするための集約型情報起点サイトとして多数のポータルサイトが開設されている。イントラネットの世界においても、ナレッジマネジメントを実現するためのツールとして企業情報ポータルを運営する企業が増えている。また、Web サービスやグリッドなどの最先端要素技術の領域でも、ユーザと直接対話するインターフェース部にポータルを採用する動きが見られるようになってきた。

今やポータルは、サービス提供者とエンドユーザとの間で中継役として各種サービス配信を行う代表的な情報ハブのデファクトスタンダードになりつつある。

この流れに伴って、2003 年後半からポータル関連技術の標準化及びオープンソース開発が進行中である。

ポータル製品間で互換性が無かったポートレット/ポートレットコンテナ実装は、2003 年 10 月の JSR168[1] 標準化によって、少なくとも Java Platform では可搬性が確保された。リモートポートレットに関する通信プロトコルの標準化とオープンソース実装も進んでいる。リモートポートレットとは、従来ポータルと同じサーバ内に存在しなければならなかったポートレットを、インターネットを介してサーバに配置できるようにする機能である。リモートポートレットの通信プロトコルは、2003 年 8 月に WSRP[2] との名称で標準化された。これに準拠したオープンソースは ASF (Apache Software Foundation) のプロジェクトの一つである WSRP4J[3] が 2003 年 9 月に開始され、現在開発の途中である。

一方、現行の WSRP 及び開発途中の WSRP4J では、リモートポートレットの動的な検索機能がなく、リモートポートレットの本格的な運用を考えた場合問題である。そこで本研究では、WSRP サービス公開/発見の仕組みに対する独自のアプローチとして、WSRP リモートポートレット検索エンジン (SERP; Search Engine for Remote Portlets) を提案し、プロトタイプ検証を行った。

2 Web Services for Remote Portlets

2.1 概要

WSRP (Web Services for Remote Portlets) は、リモートポートレットを Web サービスとして実現するためのプロトコルを規定した OASIS の国際標準仕様である。

WSRP で規定されたポートレットが提供するサービスを WSRP サービスと呼ぶが、従来のデータ指向 Web サービスに対して、WSRP サービスはユーザインタフェースを有する対話型プレゼンテーション指向 Web サービスと位置付けられる。また、WSRP ではプロデューサ (サービス提供側) とコンシューマ (サービス消費者)、及びエンドユーザ (コンシューマ利用者) の三種類のアクタが定義され、プロデューサとコンシューマ間の通信プロトコルを規定するために、プロデューサ側で統一された 4 つのインタフェースを備えることが義務付けられる。これらのインタフェースは Web サービスの記述言語である WSDL によって記述される。

2.2 WSRP サービス発見の仕組み

ところで、現行の WSRP バージョン 1.0 は、Web サービスの三フェーズ「登録」「発見」「結合」の内、プロデューサとコンシューマ間の結合仕様のみを対象としたものであり、他のフェーズについては現時点で未策定の状況にある。プロデューサの論理的な位置やそのサーバが保持する WSRP サービスの内容をコンシューマが予め知っている環境というのは極めて限定的であるため、「登録」「発見」フェーズの標準化は、今後 WSRP サービスの本格的な利用に必須の要件である。

3 WSRP リモートポートレット検索エンジン

3.1 SERP の必要性

プロデューサのエンドポイント URL が固定で、ポータル管理者がそのプロデューサをも管理しているような限定的な環境においては、リモートポートレットを検索する必要性はない。全てのサービスの状態を把握できるからである。しかし、ポータル管理者とプロデューサ管理者が異なるような、より一般的な状況においては、ポータルにリモートで接続するプロデューサやそのプロデューサが保持するポートレットの数が動的に時事刻々と変化していくため、リモートポートレットの検索機能がなければ、フレキシブルな運用が難しい。従って、SERP を開発する必要がある。

3.2 SERP の機能

SERP の機能は、「収集」「インデックス化」「検索」の三種類に分類できる。それぞれの機能は、以下の通りである。

収集 WSRP サービス検索用ロボット serpbots が WSRP プロトコルを利用して WSRP サービスのメタ情報を取

^{*}fnas-ito@complex.eng.hokudai.ac.jp
[†]札幌市北区北 14 条西 9 丁目北海道大学大学院情報科学研究科

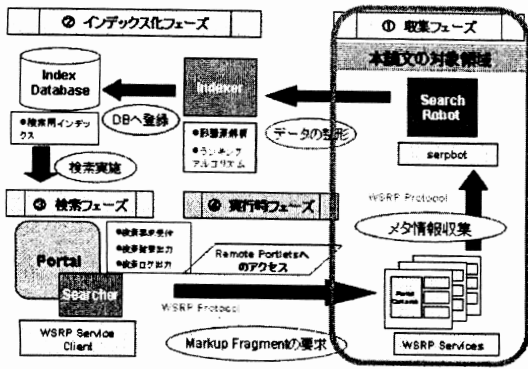


Fig. 1 SERPの全体構造

集する。

インデックス化 インデクサが収集したデータを検索用データとして加工する。この段階で、形態素解析、シソーラス作成、検索速度を速めるためのインデックス生成など、検索結果を高機能なものにするための複数の処理が行われる。

検索 サーチャが検索用データの中から、検索エンジン利用者の検索クエリに対応したデータを探し出し、検索結果として利用者に WSRP サービスのメタ情報の一覧を表示する。

4 実装

SERP のプロトタイプ実装には、WSRP4J のプロダクトを利用した。具体的な実装方法としては、まず、WSRP4J Consumer のソースコードをベースに serpbot のプロトタイプ開発を行った。serpbot は、XML で記述したプロデューサのエンドポイント URL を読み込み、順次そのサービスディスクリプションインタフェースへメタ情報を要求して、返り値として取得したメタ情報をデータベースへ格納していくという振る舞いをする。また、インデクサ及びサーチャについては、あくまで簡易実装に留めた。

Fig.2 は、SERP の実行例のログである。

```

検索結果 8個のポートレットが見つかりました。
(検索に用いたキーワード: news)
Search keywords : news

-----
ポートレットキー: 1_11
グループID:
ポートレット情報更新日時: 2004-01-31 15:58:57.236
Endpoint URL
このポートレットを所有するプロデューサの情報
マックアップインタフェースURL: http://159.254.99.229:8081/wsrp/wsrp4j/WSRPBaseService
サービス記述インタフェースURL: http://159.254.99.229:8081/wsrp/wsrp4j/WSRPServiceDescriptionService
登録インタフェースURL: http://159.254.99.229:8081/wsrp/wsrp4j/WSRPRegistrationService
ポートレット管理インタフェースURL: http://159.254.99.229:8081/wsrp/wsrp4j/WSRPPortletManagementService

ポートレットがサポートするMARKUPの種類
Markup Type
MIMEタイプ: text/html
サポートされたモード: wsrp view, wsrp edit
サポートされたWindow状態: wsrp normal, wsrp maximized, wsrp maximized
サポートされた言語: en

リソースバンドルの詳細
Resource Bundles
タイトル: Entertainment (en)
ショートタイトル: Entertainment (en)
ポートレットの特別なアプリケーション (en)
Entertainment Portlet provides entertainment information. The Portlet is a Sample Web Application (en)
ポートレットが所有するキーワード:
C_company (en), news (en), entertainment (en), movie (en), music (en)
(以下省略...)
    
```

Fig. 2 SERP の検索結果 (ログ情報)

5 考察

一階層のプロデューサについては、SERP を用いれば実時間内でリモートポートレットのメタ情報を収集できることが確認された。後は、DB のチューニングやインデクサの張り方に関する最適化手法、検索の容易さに影響を与えるランキングアルゴリズムや Web アプリケーションとしてのユーザインターフェースの改善等を、既存の Web 検索エンジンでも用いられているものを利用するなどしてさらに検討していくことで、実用に耐えうるものに作りこんでいくことが可能である。

しかし、将来のポータル像を考えた際、WSRP サービスは必ずしも一階層に限定されるものとは限らない。ポートレットのプロキシとして働くプロキシ・ポートレットを利用すれば、多階層へ拡張したサービスを提供できる。

ただし、多階層へ拡張すると現在の WSRP 仕様ではメタ情報収集に関する弊害が現れる。現在の WSRP 仕様のサービス記述インタフェースは、多階層型プロデューサを想定した仕様になっていないのである。以下の図 (Fig.3) は、この問題の解決として、現在検討中の二つの方式を示している。

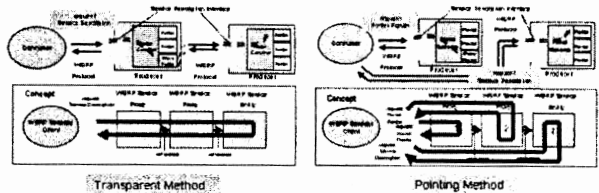


Fig. 3 透過型方式と参照型方式

6 おわりに

本稿は、現在未定義の WSRP サービス公開/発見の仕組みに対する独自のアプローチとして、SERP を提案した。SERP の機能を設計し、中でも特に収集フェーズに注力して一階層のプロデューサに対応した serpbot のプロトタイプ実装を行った。この結果、一階層のプロデューサからメタ情報を収集する仕組みの実現可能性について確認できた。さらに、現在の WSRP 仕様では多階層へ拡張する際、サービスディスクリプション仕様を拡張する必要があることが分かった。今後は、ポータル関連技術の動向に注目しつつ、SERP の本格的な開発、及びサービスディスクリプション仕様の拡張の OASIS への提案などを視野に入れていきたい。

参考文献

- [1] Alejandro Abdelnur, Stefan Hepper, Java Portlet Specification version 1.0, Sun Microsystems, 2003
- [2] Alan Kropp, Carsten Leue, Rich Thompson, Committee Specification 1.0: OASIS Web Services for Remote Portlets, OASIS, 2003.
- [3] Welcome to WSRP4J, <http://ws.apache.org/wsrp4j/>