

ISDL レポートシステムの提案

大西 祥代

1 はじめに

現在、我々の研究室では、学生が研究報告、文献調査などのレポートを HTML 化し Web 上で公開しており、これらを ISDL レポートと呼んでいる。ISDL レポートの公開は、我々の研究活動を広く外部に知ってもらうことを目的としている。これまでに、約 1300 本ものレポートが公開され、アクセスログを解析した結果、外部からの一定アクセス数が確認でき、研究室の内部外部共に有用なデータとして活用されていることがわかった。しかし、現状では ISDL レポートを閲覧する十分な仕組みが整っておらず、システムに改善の余地が多く残されている。そこで、ISDL レポートの公開方法を見直し、ISDL レポートへのアクセス数を向上させるシステムを作成する。

2 ISDL レポートシステムの現状

現在、ISDL レポートを集めたポータルサイトが存在するが、以下の点で改善が求められている。

- 初めてポータルサイトを訪れたユーザにとってサイト構成が分かりづらい。
- どのような研究レポートがあるのか把握できない。
- レポートの検索方法が学生の名前によるものだけである。

また、ほとんどの ISDL レポートはレポート間にリンクがなく独立して存在しており、関連するレポート同士を結びつける仕組みが存在しない。これは多くのレポートを公開しているにも関わらず、ユーザがアクセスしたレポート以外のレポートを閲覧する機会を失っている。さらに、ユーザは関連したレポートをユーザ自身の手で探さなければならず、効率的な情報検索ができない。

3 提案する ISDL レポートシステム

2 章で述べた現在の ISDL レポートシステムを改善し、ユーザにより多くのレポートを閲覧してもらい、効率的に情報検索を行うことのできる ISDL レポートシステムを提案する。提案システムでは、ISDL レポートポータルサイトの整備により、ユーザにやさしいインタフェースを提供する。そして、レポート同士のリンクを自動的に生成し、ユーザにより多くの情報を提供する。

3.1 ISDL レポートポータルサイト

現在の ISDL レポートポータルサイトでは、初めてのサイト訪問者にとってサイト構成が分かりづらい。また、検索方法が学生の名前だけなので、研究室内部の者にとってもレポートの検索が難しい。そこで、以下の点を加えた新たな ISDL レポートポータルサイトを作成する。

- 初めてのサイト訪問者に、研究室の概要や研究分野を理解できるよう、研究マップを作成する。
- フリーキーワード、著者、研究グループ、研究室を代表するキーワードによる検索機能を付加する。
- 週間アクセスランキングを表示する。
- 新着レポートを表示する。

これらの機能を付加したポータルサイトは Fig.1 に示す構成である。

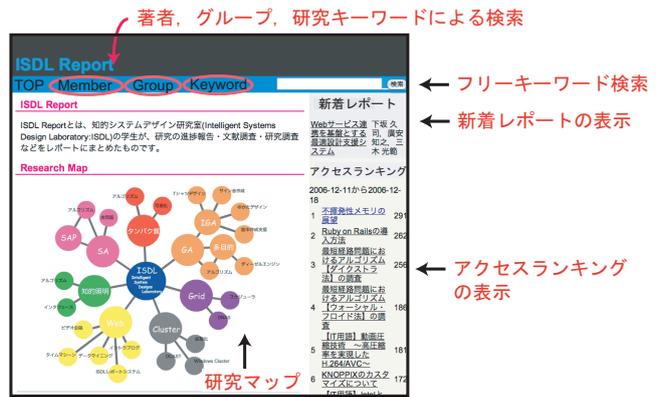


Fig.1 新しい ISDL レポートポータルサイト (出典 : 自作)

3.2 レポートの間リンク

現在の ISDL レポートシステムでは、レポート間にリンクがないため、一つのレポートの閲覧から他のレポートにアクセスして情報を得ることがなかった。そこで、レポート間の関連度を計算し、関連するレポートに自動的にリンクを張ることで、一つのレポートの閲覧でより多くの情報を提供する。本システムでは、レポート全文を用いてレポートの関連を調べる。関連度計算には GETA¹⁾ (Generic Engine for Transposable Association) を用いる。

3.2.1 GETA

GETA は日立製作所などのグループで開発された、検索する文書の単語の出現頻度を行列で表現したのに対して、文書間の類似度を高速計算するツールである。単語の出現頻度を求めるためには、文書に対して形態素解析を行う必要がある。本システムでは形態素解析に MeCab²⁾ を用いる。

GETA では関連度を計算する際、文書中の単語の重要度を用いる。重要度は以下に示す tf と idf の積である、tf * idf 値をもとに計算する。

- tf(Term Frequency) : 1 つの文書に何回も出現する単語は重要語

- df(Inverse Document Frequency) : 多数の文書に出現する単語は重要語でない

3.2.2 レポート間のリンク作成手順

レポート間にリンクを張るための手順を以下に示す．

1. MeCab を用いて全レポートの形態素解析を行う．
2. 各レポートの単語の出現頻度を数える．
3. GETA により単語出現頻度のバイナリファイルを作成する．
4. 1つのレポートをクエリとして，単語出現頻度のバイナリファイルとクエリのマッチングを行う．このとき $tf * idf$ が利用される．
5. 重要度の高いものから順に出力する．
6. 重要度の高いレポートにリンクを張る．
7. 全レポートに対して 4~6 を繰り返す

以上の手順により，関連するレポートへのリンクを作成することができる．Fig.2 にリンク作成の例を示す．各レポートについて形態素解析を行い，各レポートを行，出現単語を列とする行列を作成し，単語の出現頻度を数える．Fig.2 の行列をもとに，GETA がレポート間の関連度を計算する．例では，ReportA と ReportB が「Web」，「アプリケーションフレームワーク」という単語の出現頻度が高いので関連が高くなる．レポート数が増すと，単語出現頻度の行列が膨大な大きさになるが GETA により高速に計算することが可能である．

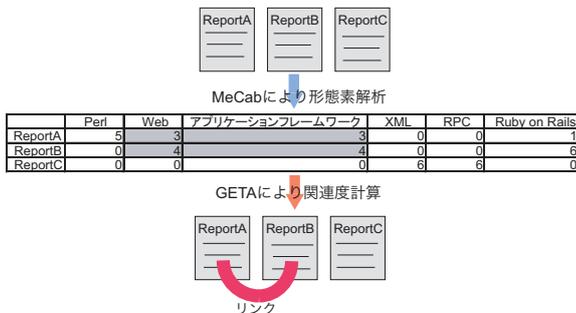


Fig.2 レポート同士のリンク (出典：自作)

Fig.3 は ISDL レポートに関連レポートの情報を付加したものである．このように，全てのレポートの中に関連レポートを提示することができれば，レポートのアクセス向上が期待でき，ユーザにとって利便性の高いものになる．



Fig.3 関連レポートの提示 (出典：自作)

4 今後の課題

4.1 新しい ISDL レポートシステムへ導入による効果の調査

新しい ISDL レポートシステムでは，ポータルサイトの整備を行い，レポート間のリンクを作成した．今後はポータルサイトの利便性についての調査や，リンク作成によるアクセス数の影響の調査を行う必要がある．

4.2 ユーザ参加型コンテンツへの発展

現在の ISDL レポートは著者がレポートを公開するだけで，レポートを閲覧したユーザがレポートに対して意見を残すことができない．レポートに対して興味を持って，現在は研究室の連絡先を調べてメールを送る方法しかなく，気軽にレポートの著者と閲覧したユーザがコミュニケーションをとることができない．そこでレポートに対してコメント機能やトラックバック機能を付加することにより，レポートの著者と閲覧したユーザとのコミュニケーションの促進を図る．これにより，外部の研究者からフィードバックを得られる機会が増し，研究意欲の向上や，さらなる研究の発展が期待される．またトラックバック機能により，外部のブログと繋がることで，本システムに対する被参照情報を得ることができ，参照元の情報を含んだ検索が可能となる．このように，レポートの著者だけでなく閲覧するユーザも参加できる，ユーザ参加型の ISDL レポートシステムを目指す．システムの全体像を Fig.4 に示す．

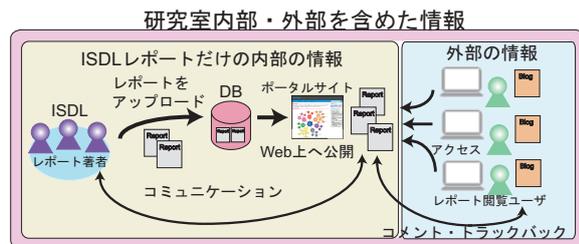


Fig.4 ユーザ参加型の ISDL レポートシステム (出典：自作)

5 まとめ

本報告では，新しく ISDL レポートポータルサイトを構築し，レポート間のリンクを作成した．これによりレポートを閲覧する仕組みを整え，ユーザにより多くのレポートを提供することが可能となった．今後は新しいシステムの効果を調査する必要がある．また，現在のシステムにコメント機能やトラックバック機能を付加することで，ユーザ参加型のシステムを目指す．

参考文献

- 1) <http://geta.ex.nii.ac.jp/> 国立情報学研究所，高野明彦 日立製作所，西岡真吾
- 2) <http://mecab.sourceforge.jp/> 京都大学情報学研究所・日本電信電話株式会社コミュニケーション科学基礎研究所共同研究