

# ISDL 文書システム

## System for ISDL reports

箕輪 祐子  
Yuko MINOWA

**Abstract:** Our laboratory's research reports are made public on the Internet. By analyzing data from the access logs of visitors who use these reports, we are able to come up with new fields and topics of research.

## 1 研究目的

本研究室では、ISDL 文書としてウェブ上で研究テーマに沿ったレポートを公開しており、レポートのテーマは自らが決定し発信している。

これらのレポートを利用する訪問者からのニーズを知ることによって、新しい研究分野やテーマを発見できる可能性がある。本研究では、ISDL 文書のログをデータマイニングすることで必要な研究やレポートの内容を考える。そのためにインターフェイスとコンテンツを充実させ、訪問者の意思に準じたデータを分析することを目的とする。

## 2 研究内容

### 2.1 インターフェイス

ISDL レポートのインターフェイスは検索とリンクであり、これらを充実させることが基盤となる。

現在、訪問者が閲覧する ISDL レポートは、検索エンジンの条件検索において見つけられたものが 3 割から 5 割を占める。しかし、各レポートには他のレポートへのリンクが少なく、レポート間の繋がりが薄いため、訪問者は検索を以て対象のレポートを見るだけで終わってしまうことが多い。よってまず ISDL レポートのインターフェイス部分の改善が必要である。

#### 2.1.1 検索エンジンへの対応

英訳をすることで、英語でのアクセス数を上げる。この英訳では、研究に関わるキーワードの英語変換をこちらで指定できる必要がある。

#### 2.1.2 関連ページへのリンク

閲覧するレポートに関連レポートへのリンクが含まれると、訪問者はリンクを辿り様々なレポートを見るのが可能になり、更なるアクセスに期待できる。

また、そのレポートを辿る一連の流れから何を探しているか、またレポート同士の新たな繋がりを通じて研究テーマを発見することが望めれば、後述のニーズに繋がると考える。

### 2.1.3 トップページ

レポートが格納されているトップページの使い勝手をよくすることで、トップページからのアクセスをよりし易くする。検索や、レポートのタイトルだけでは不明瞭であった内容をアブストラクトの表示とかけ合わせることで明確にし、機能を増やす。

## 2.2 コンテンツ

インターフェイスを整えた上で、利便性を考えたコンテンツを増やすことでさらなる充実を図る。

### 2.2.1 単語帳

キーワードの意味を自動リンクで確認できる単語帳を作成することでレポートの理解を深める。

## 2.3 ニーズ

### 2.3.1 アクセスログのデータマイニング

ウェブサイトへアクセスしてくるクライアントは、さまざまな情報を送ってくる。これらの情報を解析することで訪問者のニーズをよりよく知ることができる。解析できる項目は様々であるが、代表的なものとして以下のものが挙げられる。

- 検索エンジン
- 検索用語
- リンク元統計
- アクセス数  
時系列やページ毎による傾向や変動
- リピータ調査  
回数・前回訪問からの時間
- IP・ブラウザ  
ドメイン・OS・ブラウザ・国/言語
- ディスプレイ

これらの個別の情報を複合して調べることで、色々な角度からデータを収集することができる。同じ IP がどのレポートにアクセスしたか、レポートの遷移状況やその時間的間隔、リピート率などの複合統計を出し、さらにマイニングする。

### 2.3.2 サイト内検索

例えば、SA で検索した訪問者が GA という単語をレポート内に見つける。この GA について調べたいと訪問者が考えたとき、同じ ISDL 文書内の GA のレポートを見つける機能があれば利便性は増し、さらなるニーズを探ることができる。

また、検索エンジンからの検索では、ウェブページに存在する用語でしかレポートにたどり着かないため、ウェブサイトが存在しない単語は知りえない。新しい単語を知る他にも [SA Perl] など、新しい組み合わせを発見することを期待する。

## 3 今後の課題

インターフェイスを充実させることが第一であり、自動英訳や関連ページへのリンクなど、チェッカーを使用し全ての ISDL レポートに対応させる。その後、ログの解析を行い、結果をデータマイニングするためにもその方法を検討する必要がある。