



Date : 13/07/2006

Title: Shopping for current day news in library database. Is it possible?

**Ilona Dukure
Anita Goldberga**

**ilona.dukure@lnb.lv
anita.goldberga@lnb.lv**

Meeting:	123 Cataloguing
Simultaneous Interpretation:	Yes
<i>WORLD LIBRARY AND INFORMATION CONGRESS: 72ND IFLA GENERAL CONFERENCE AND COUNCIL</i>	
20-24 August 2006, Seoul, Korea	
http://www.ifla.org/IV/ifla72/index.htm	

今日のニュースを買って読もう。

——図書館データベースにおける可能性

訳者注：原文中の画像は、技術的な理由により、赤い線で表示されています。画像については原文を参照してください。

概要 ラトビア国立図書館が、IT 企業である Lursoft IT 社と共同して行った事業「ラトビア新聞記事配信モデル：Latvia's Model for Periodicals' Content Distribution」について報告します。この計画は、新聞記事の全文テキストへのアクセスを高めることを目指しています。本報告では、国立図書館がどのように電子的資源の書誌コントロールを行うのか、その潜在的な可能性について論じます。また、電子的資源からメタデータを自動生成して図書館のデータベースに取り込む可能性についても報告します。利用者や事業のパートナー企業にとってのメリット、懸案となっている問題点についての指摘も行います。

ラトビアからの報告です。ラトビアはエストニアとリトアニアにはさまれたバルト海沿岸の国です。西にはスウェーデン、東はロシアが隣国としてあげられます。

インターネット資源をメタデータとして組織化するには、以下の方法が考えられると思います。

1. 既存の図書館情報システム（MARCフォーマットであることが一般的）を利用してインターネット資源を記述する。
2. デジタル資源の書誌コントロールのために別個の体系を構築するか、応用する。（ダブリンコア、または、その他の簡易な構造のフォーマット）
3. 出版者作成のメタデータを利用する。
4. 情報資源からメタデータの自動生成を試みる。
5. 記述対象である情報資源の多様性を考慮し、上記のすべての方法を組み合わせて利用する。
6. ウェブ上でアーカイヴされ利用可能な情報資源に対しての索引を作成し、出版する。もしくはそれに類似の手法。
7. ポータルサイト、またはGoogleのようなサーチエンジン（ラトビアでは、Siets）で検索できるように図書館作成の構造化されたメタデータを活用する。

ラトビア国立図書館（National Library of Latvia: NLL）は、デジタルボーン資源の書誌コントロールに、1枚目のスライドで紹介した方法のうち3つをすでに実施しています。まず、スライドで1番目に紹介した伝統的な手法です。次に、（スライドでは3番目にあたりますが、）出版者と手を結び、これから出版予定の出版物へリモートアクセスでISBN付与（将来的にはISSN付与も）を行う共同事業です。これは現在継続中であり、インターネットを利用して、これら出版物のドラフト入力段階のメタデータを全国書誌データベースに供給することも行っています。最後に、スライドの7番目に挙げた事例として、欧州の他の国立図書館と行っているTEL-ME-MORプロジェクトの一環として行っているものがあります。

出版者の動向や活動を視野に入れなくては、われわれはインターネットを（書誌コントロールの観点からは）網羅的に取り扱うことができません。ラトビアでも活動している電子出版者はありますが、図書の出版者であるケースは多くありません。そこで、最初の段階では、逐次刊行物に焦点を絞ることにしました。今日は、上記スライドの4番目の手法についてラトビアの事例を詳しくお話したいと思います。名づけて、**ラトビア新聞記事配信モデル**です。

2005年、ラトビア国立図書館（NLL）は、ラトビア記事索引データベース（NLL National Analytics Database:）内を検索すると、新聞、雑誌、ジャーナルの全文テキストへインターネット経由でアクセスできることを目指した共同事業を開始しました。このデータベースには、1988年から2005年の間にラトビアで刊行された新聞、雑誌、ジャーナルの記事のメタ情報（約150万レコード）が収められており、Aleph500システム（version 505.14.2）により、インターネット上での表示を行っています。（http://www.lnb.lv/eng/db/nat_bibl.htm）

ラトビア記事索引データベースは1996年にスタートしました。2002年以降は、MARC21を応用したAleph500という図書館情報システムによってレコードを作成しています。旧ローカルシステム作成のレコード（66,6700件）も、Aleph500へ移行されています。

ラトビア記事索引データベースは、著者名、タイトル、記事の出处（逐次刊行物のタイトル）、キーワード、論文中に登場する個人名などで検索できます。ラトビア国内、国外を問わず広く利用されていて、ラトビア語、ロシア語、英語などで刊行されたラトビアの最新の逐次刊行物からの情報を提供しています。

ラトビアでは年間600タイトルの逐次刊行物が出版されていますが、その3分の2がラトビア記事索引データベースに収録されています。漫画、事業案内書、クロスワードパズル、テレビ・ラジオ番組表のようなものの記事採録は行っていません。

ラトビア記事索引データベースから全文テキストへのアクセスは早くから考えられていたのですが、2005年5月にその実施を開始しました。

”文化情報システム機構”は、ラトビアの国レベルの図書館共同情報システムの運営を行っている機関ですが、この機構により、2004年、書誌データ/全文テキスト間にリンクを生成するための入札が行われ、その結果、Lursoft IT社と契約を取り交わし、ラトビア記事索引データベースから全文テキストへのアクセスを提供することになりました。

われわれの共同開発パートナーとなった Lursoft IT 社(<http://www.news.lv>) は、ラトビア国内で刊行された電子新聞、電子ジャーナルの全文テキストのアーカイブを行っている会社です。アーカイブには、首都圏および地方の主要な刊行物のデジタルコピーが保存され、その収集は Lursoft IT 社と各出版者との契約に従って法的に行われています。270 万 7 千件以上の記事の全文テキストが Lursoft IT 新聞ライブラリにおいて利用でき、その数は毎日約 1000 件増加していきます。

2006 年現在、Lursoft 電子新聞ライブラリを通じて、43 の新聞および通信社の論説記事へアクセスすることができます。表、図やグラフの表示はできませんが、全文テキストを読むことができます。また、写真、事業案内、広告のデータは入っていません。

この開発に当たり、電子出版物の利点を最大限生かすために、Lursoft IT 社は URL だけではなく新聞の紙面から得られる他のデータも提供することにしました。

ラトビア国立図書館 (NLL) の方は、フォーマット、レイアウト、コード化情報について、送信されるべき記述の必須要素を定義しました。

2005 年 5 月、Lursoft IT 社の新聞記事メタデータ自動配信が開始されました。メタデータは、新聞発行後に、日次で自動的に紙面より抽出され、出版者から Lursoft IT 社に送信されてきます。

1993 年以降 Lursoft IT 社は、協定を結んだ発行者 (<http://www.news.lv> 左欄のリスト参照) から、新聞の全紙面を受け取っています。記事は紙面から機械的に (一部、人間が補正をしますが) 抽出し、Lursoft IT 社が特別に開発したソフトウェアを使って、構造化された XML データベースのエントリに変換します。そこから、専用ツールを使用し、データの書誌データ要素部分を変換し、ラトビア国立図書館 (NLL) へ送信します。

XML データベース管理システムなど Lursoft IT 社が開発した技術(詳細は、<http://www.siets.biz> を参照)は、内外を問わずニュースデータベースや公文書データベースに利用されています (<http://www.siets.biz/solutions/usecases/> と <http://www.siets.biz/company/> を参照)。

Lursoft IT 社から送られてくるデータには、以下の記述が含まれています。記事の著者名 (著者が存在する場合)、記事のタイトル、新聞のタイトル、巻次・年月次、各記事特有の識別子。その日その日の適切な出版記事の記述は、午前 3 時まで、あらかじめ定められたフォーマット (Aleph Sequential--図書館情報システムへのダウンロード用) で送付されてきます。

情報管理システム Aleph500 は、文化情報システム機構の管理下にあり、同機構が、毎日受け取ったファイルを Aleph500 のシステムに流し込みます。流し込む先として、独立したデータベースを作成しました。受け取ったファイルのレコードすべてをラトビア記事索引データベースに流し込むわけではないからです。首都圏で出版されたもののデータの約半数、地方出版紙の 4 分の 1 から 6 分の 1 がラトビア記事索引データベースの目録対象になります。研究報告、経済、政治、歴史、地方研究などの重要度が高く価値のある記事だけが選択されて、目録が作成されます。

上の画面が、Lursoft IT から Aleph500 書誌作成機能に取り込まれた構成レベル書誌の画面です。

レコードには必須の固定長のコード化情報があります。データベースへのデータ投入の際、これらの情報が存在しないと投入が行えません。さらに、Lursoft IT に依頼して、目録情報作成機関、言語、全国書誌識別記号、URL などのフィールドを追加してもらうようにしました。これらのフィールドがあらかじめ存在することによって、目録作業者は手作業でフィールドの追加をする必要がなくなり、作業の省力化になります。

ラトビア国立図書館の専門スタッフが、これらのデータをラトビア記事索引データベースに登録するとともに、MARC21 の要件にしたがって、受け取ったデータの質を向上させます。

上のスライドでは、カタログによって処理が終わったレコードをご覧に出来ます。UDC、注記、記述情報の追加、標目、付属資料の情報があります。

さて、このラトビア記事索引データベースはラトビア国立図書館のサーバに、リンクする全文テキストファイルは Lursoft IT サーバに存在します。

インターネットからユーザが検索を行うと、ラトビア記事索引データベースのレコードから記事固有の識別子（フィールド 856）をキーとして対応する全文テキストへのリンクがたどれます。ラトビア記事索引データベースへのアクセスは無料ですが、全文テキストへのアクセスはそれぞれの新聞の発行者との取り決めにより異なります（アクセスは許可制）。Lursoft IT デジタルアーカイブは合法的な収集を行っているため、著作権保護の観点からも問題はありません。

では、実際にどのように全文テキストまでアクセスするのかを、NLL のホームページへからみてみましょう。こちらです (http://www.lnb.lv/eng/db/nat_bibl.htm)。"Database" のメニューから、"National Analytics database" (ラトビア記事索引データベース) を選択します。

検索項目を選択します（例えば、タイトル）。ここでは、ホッケーについて検索してみましょう。62件のレコードがヒットしました。

選択したレコードを開くと記事の所在を示すリンク情報があります。

このリンクをクリックすると、著作権情報のウィンドウが現れます。

認証されると、検索した記事の全文テキストが閲覧できます。

この共同事業は現在も開発が継続中です。現在、新聞全文テキストアーカイブと1993年から2004年分NADレコード40万件の間に、遡及でリンク生成を行っています。

2005年からEUの援助を受けて開始されたプロジェクトであるTEL-ME-MORのフレームワークを利用しても、欧州の国立図書館のポータルからもラトビア記事索引データベースにアクセスすることができます。

問題点：

- Lursoft IT社からの日時更新が1日2回行われてしまうことがあります。
- すべての記事に全文テキストへのリンクが張られるわけではありません。記事によっては発行者が公共の利用に適さないと判断することがあるからです。
- 当たり前のことですが、ユーザがすべてのサービスを無料であると誤解してしまうという問題があります。

しかし、これらの問題点はプロジェクトを大幅に軌道修正させるにはあたらないものです。

成果:

- 「発行ーメタデータ作成ーデータへのアクセス」が同日内に可能
-

- **NLL** は、
 - 当日の電子記事のドラフト状態のメタデータを受け取ります。
 - データの質を充実させます。タイトル、著者名を手入力で修正する際に発生する文法エラーは、自動的に修正されます。
 - **出版者**は、どのような記事、どの執筆者へのアクセスが多いかについて、間違いのない統計を入手できます。
 - **出版者と Lursoft IT 社**は新規ユーザを開拓できます。
 - **利用者**は、記事の全文テキストからリンク先の情報（例：記事中に引用されている個人名）へ切り替えることができます。商業データベースのリンク先情報や他の正式登録簿との接続が、同一インタフェースで可能です。
 - **利用者**は、24 時間オープンの情報メーション・ショップとして利用できます。
-

12. 04. 2006

Ilona Dukure ilona.dukure@lnb.lv

Anita Goldberga anita.goldberga@lnb.lv

Translated by: Inahama Minoru