Five Years of

Tax Map:

A Topic Maps Application at the U.S. Internal Revenue Service: Lessons learned

Michel Biezunski, Infoloom

Topic Maps 2007

Oslo, Norway, March 21st, 2007

Contents

- Project Motivation and History
- TaxMap as a Bottom-Up Application
- TaxMap and the Topic Maps Model

Original Motivation

- Improvement needed on answers from IRS to taxpayers' questions.
- Congressional hearing determined quality level of help offered to taxpayers was not satisfactory (in 2000).
- The amount of information is overwhelming for taxpayers' assistors.
- Need research tool for taxpayers' assistors organized on topic navigation.

Project History

- 2001: Prototype. Since 2002: Production
  - First prototype : 8 *Publications*
    - Extended to 33 publications, then all IRS publications (150+)
  - Augmented with the *Tax Law FAQs* (several hundreds)
  - Extended with the *TeleTax Topics* (~100)
  - Extended with the *Forms* and *Instructions* (700+)
  - Many improvements and extensions over the years.
- Technology
  - Technology used: Topic Map Loom, created by Michel Biezunski, used since 1996 for Conference Proceedings, Encyclopedia, Book Publishing and other applications.
  - Team: with Coolheads Consulting and Plexus Scientific.

Tax Map Today

- Research tool available on IRS Intranet for taxpayers' assistors in call centers
- Available in 2 Cd-Roms for taxpayers
  - Small Business Resource Guide
  - Tax Products CD (Publication 1796)
    - Order from http://www.irs.gov
- Available on the Web from various unofficial sources (Google "IRS Tax Map")

Initial Requirements

- Should enable research by subject.
- Should be fast and easy to use.
- Should meet the needs of the visually impaired.
- Should not add work to IRS authors.
- Should leverage existing SGML/XML investments.
- Should enable constant updates.

Challenge: Research By Subject

- Topic Maps standard provides a solution.
- Model enables multiple names to be used for the same subject.
- Implementation enables navigation to occurrences, between occurrences, from occurrences and to and between related topics.

Challenge: Fast and Easy to Use

- Fast to run:
  - Load on server minimized: product is a set of pre-created HTML files.
  - Search engine on the client only on topic names.
- Easy to use:
  - Use Web browser Interface.
  - Topic Map model used for preresolved queries:
    - Topical index (List of topic names). Focus on key topics.
    - Topic pages with occurrences and relations.
    - Cross-occurrence navigation.
  - Five minutes' training sufficient.

Challenge: Don't touch anything!

- Authors will not change the way they work.
- Topic Maps model considered as an overlay
  - outside the information sources.
- Application designed as bottom-up.

Challenge: Use of SGML/XML Work

- IRS pioneered SGML (since before 1985).
- Four different authoring communities and processes:
  - *Publications* in SGML or PDF.
    - 2 different DTDs.
    - To be converted into XML.
  - *Forms* in PDF, *Instructions* in XML and/or PDF.
  - *TeleTax Topics* in XML.
  - *FAQs* in XML.
- 5 different DTD/schemas used concurrently.
- TaxMap cannot impact DTDs and Schemas.

Challenge: Constantly updated

- Batch process:
  - Inputs:
    - IRS documents undergo substantive updates annually.
    - Documents to be included in TaxMap vary.
    - Expert knowledge base maintained separately.
    - Configuration files (including style sheets) get improved.
  - Outputs:
    - Several variants (Intranet, Cds) of navigable topic maps in HTML.
    - A number of reports, including an audited version.
    - XTM (used to be part of the output).
- Produced about once a week.

A Bottom-Up Approach

- Extract information from sources
  - Topic names from content of certain elements.
- Tweak results
  - Combination of automatic and manual processes.

Automatic Rules

- Extracting topics from markup
  - Content of certain elements make topic names.
  - Eliminate prepositions from names.
  - Assimilate plural forms with singular forms.
  - Acronyms (eliminate when present).
  - Delete topics following certain patterns.
- Acquisition of topics from IRS product database
  - Eliminate relations with products not included, etc.
- Creation of relations between topics
  - E.g., “Containment rule”.

Human Input

- Renaming topics (new name may or may not already exist).
- Merging (two names become “synonyms”, assigned to the same topic. All other properties merge).
- Deleting topic by name. This will delete all its other properties.
- Typing. Assigning a topic type.
- Dividing. Topic name indicates more than one topic.

Initial Model: Topic Maps Constructs In

- Topics
  - Names: Extracted from content of certain XML/SGML elements.
- Occurrences
  - From within publications, FAQs
  - Occurrence type: (context) document title + section header
  - Occurrence scope: type of document in which occurrence is found.
- Associations
  - “Related” semantic.

Initial Model: Topic Maps Constructs Out

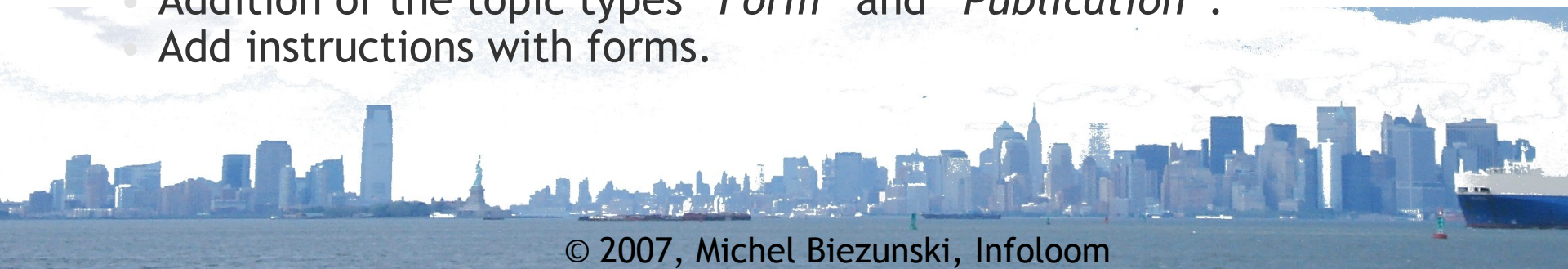
- No topic type.
- No association type.
- No association role.
- No scope for names.
- No scope for associations.
- No variants.
- No resource reference indicators (everything is a subject indicator).

The Index Story

- Each publication has a back-of-the-book index.
- Naturally, we started by extracting topics from index markup.
- Integration became a problem. Discussions started on how to improve the indexes.
- Professional indexers came into play. They decided to harmonize terms but also increased significantly the number of indexed terms.
- The resulting integrated index was unmanageable.
- We altered TaxMap: we decided to extract section headers instead of index terms.

Topic Map Model. Evolution

- The index was too big
  - New topic type: “Key Topic” for topics that would appear in the indexes.
- Subtopics inconsistent:
  - Subtopic relationship type removed.
  - Later, pub indexes ignored.
- Topic page needed better organization:
  - Occurrence scope: Type of document in which occurrences are found. E.g. “*Forms*”, “*Publications*”, “*TeleTax*”, “*FAQ*”
- Forms and Pubs were not findable:
  - *Publications* and *Forms* considered both as occurrences and as topics.
  - Addition of the topic types “*Form*” and “*Publication*”.
  - Add instructions with forms.



Improvements of Tax Map

- Consistency with IRS Product Database.
- Semi-annual workshops with Tax Experts.
  - Maintenance of Map Integrity

Current Rendered version

File Edit View Go Bookmarks Tools Help

TAX PRODUCTS CD
IRS TAX MAP 2004

Search All Topics:

GO

Search Help
Navigation Help

Main Topics
A B C D E F G H I
J K L M N O P Q R
S T U V W X Y Z #

Forms
Publications

Comments
About Tax Map

Publication 17

Your Federal Income Tax

For use in Preparing 2004 Returns

Introduction, Useful Items and Important Changes

Part 1
The Income Tax Return

Chapter 1
Filing Information

Do I Have To File a Return?

Individuals—In General
Dependents
Children Under Age 14
Self-Employed Persons
Aliens
Who Should File

Which Form Should I Use?

Form 1040EZ
Form 1040A
Form 1040

Does My Return Have To Be on Paper?

IRS

When Do I Have To File?

Extensions of Time To File

Inside Pub 17

≺ **Previous Page:** Publication 17 - Your Federal Income Tax - Figuring Gain or Loss
≻ **Next Page:** Publication 17 - Your Federal Income Tax - Business Use or Rental of Home
Use ≺ ≻ to find additional instances of index items.

≺ Sale of Home ≻

Excluding the Gain

You may qualify to exclude from your income all or part of any gain from the sale of your main home. This means that, if you qualify, you will not have to pay tax on the gain up to the limit described under *Maximum Exclusion*, next. To qualify, you must meet the ownership and use tests described later.

You can choose not to take the exclusion by including the gain from the sale in your gross income on your tax return for the year of the sale.

≺ Maximum Exclusion ≻

Maximum Exclusion

You can exclude up to $250,000 of the gain on the sale of your main home if all of the following are true.

1. You meet the ownership test.
2. You meet the use test.
3. During the 2-year period ending on the date of the sale, you did not exclude gain from the sale of another home.

You can exclude up to $500,000 of the gain on the sale of your main home if all of the following are true.

GO

arch Help
ation Help

ain Topics
D E F G H I
N O P Q R
W X Y Z #

Forms
blications

Comments
t Tax Map

TOPIC: Sale of Home

Synonym: Main Home Sale
Synonym: Residence, sale
Synonym: Selling Your Home

Forms Publications

Forms and Instructions

Form 1040 (Schedule D) Capital Gains and Losses MORE

1040 (Schedule D) Instructions

Form 1041 (Schedule D) Capital Gains and Losses MORE

Publications

Publication 523 Selling Your Home PDF MORE

Links Inside Publications

Publication 3 - Armed Forces' Tax Guide - Sale of Home

Rules That Provided for Postponing Gain Under the rules of this section, you ...

Publication 3 - Armed Forces' Tax Guide - Sale of Home

Sale of Home Current tax rules that apply when you sell your main home differ from ...

Publication 17 - Your Federal Income Tax - Other Income

Sale of home. You may be able to exclude from income all or part of any gain ...

Related Topic Links

- Capital Gain
- Capital Gains and Losses
- Depreciation, Recapture
- Home
- Main Home
- Postponed gain from sale of home.
- Rules for Sales Before May 7, 1997
- Sale
- Sale or Exchange of Home
- Tax Benefits When Selling Your Home
- Trading homes

The Need to Audit the Process

- TaxMap results from both automated processes and human input.
- Customer demanded to know:
  - Where does this topic name come from?
  - Why does this relation between topics exist?
  - What happened to this topic that isn't here any more?
  - Etc.

Requirements

- Auditability.
- Measure consistency improvement.
- Support collaboration of experts.
- Assessment of options for automatic processing.
- Maintenance of the topic network.

Auditing Methodology

- Record processes of creating TaxMap, both manual and automatic.

Lessons learned: Topic Maps Advantages

- Unobtrusiveness:
  - Can be applied to existing information without changing it.
- Extensibility:
  - More information can be incorporated at any time.
- Flexibility:
  - No need to be right the first time. The approach can be changed without undue expense (see the Index story).

Lessons Learned: Current Issues

- Missing features:
  - Privileged name (name type)
- Distinctions sometimes artificial
  - Between occurrences and associations (distinction doesn't exist for end users of TaxMap)
- Interchange capabilities not used (yet ?)
- Maintenance
  - Bottom-up application needs constant maintenance.
  - Maintenance needs to be thought about independently of the TM model.
- Auditability: the next challenge
  - Needs finer granularity
  - Process-oriented.