IP to VIVO XSL Information

Overview

The definition of attributes from IP and their transformation to elements in the VIVO ontology's schema.

Development Documents

XSLT Tutorial - http://www.w3schools.com/xsl/xsl_languages.asp

Xpath Tutorial - http://www.w3schools.com/xpath/default.asp

XSL-FO Tutorial - http://www.w3schools.com/xslfo/default.asp

IP to VIVO

Element/

Attribute | Name | Description | VIVO | Equivalence in RDF | Appears In |

E	Technology	Top Level element in All-Technology and contains one technology	V	RDF Element `<rfd:Description>`	Main Citation
E	Type	There are 4 types of technology: Technology, Research Tool, Material, and Innovation	V	`<rdf:Type ctsaip:Technology>`	

```
<rdf:Type ctsaip:Research Tool>
```

```
<rdf:Type ctsaip:Material>
```

```
<rdf:Type ctsaip:Innovation>
```

| Main Citation |

E	Title	The title of the technology	V	`<rdfs:label>`	Main Citation
E	Summary	The abstract of the technology	V	`<bibo:abstract>`	Main Citation
E	Description	The description of the technology	V	`<core:description>`	
E	Advantage	The advantage(s) of the technology	V	`<ctsaip:advantages>`	Main Citation
E	Institution-tech-id	The unique identifier of the organization	V	`<ctsaip:internalCaseNo>`	Main Citation
E	Institution	The name of the organization that sponsors the technology	V	`<rdf:Description><rdf:Type foaf:Organization>`	Organization

E	Contact-name	The name of the main contact person	V	`<rdf:Description><rdf:Type foaf:Person>`	Person
E	Contact-email	The email of the contact person	V	`<core:email>`	Person
E	Inventor-first-name	The first name of the inventor	V	`<foaf:firstName>`	
E	Inventor-last-name	The surname of the inventor	V	`<foaf:lastName>`	
E	Institution-link	The website of the institution	V	`<core:webpage>`	Main Citation
E	Keywords	The keywords used to describe the technology	V	`<core:freeTextKeyword>`	
E	Disease	The diseases that are of relevant to the technology	N/A		
E	Status	The status of the technology	V	`<bibo:status>`	Main Citation
E	Ctsaip-link	The website of the technology	V	`<core:webpage>`	Main Citation

Data Parsing

Universities that require parsing for *inventors* and/or *advantages* data.

- *(Partially) Implemented*. Detail in later sections.

University	Parse from <tag> to <tag> (Keywords)
• Columbia University *	<description> <advantage> (Advantages) ***<description> <inventor-first-name><inventor-last-name>*** |
• Cornell University *	<description> <advantage> (Technical Merits)
• Emory University *	***<inventor-first-name> <inventor-last-name>*** <summary> <advantage>(Highlights) |
• Harvard University *	<summary> <advantage>(INNOVATIONS & ADVANTAGES)
• Medical University of South Carolina *	<summary> <advantage>(Advantages) ***<inventor-first-name> <inventor-last-name>*** |
• NIH *	***<inventor-first-name> <inventor-last-name>(, and)***

• Tufts University *	***\<summary\>*** ***\<inventor-first-name\>\<inventor-last-name\>***
• University of Cincinnati *	\<summary\> \<advantage\>(Advantages)
• University of Michigan *	\<description\> \<advantage\>(Advantages)
• Vanderbilt *	***\<summary\>*** ***\<inventor-first-name\> \<inventor-last-name\>***

Summary

- Columbia University
  The \<description\> tag holds data for the inventors and the advantages. The inventors data should be parsed into the \<inventor-first-name\> and \<inventor-last-name\> tags and the text under the Advantages heading should be parsed into the \<advantage\> tag.

- Cornell University
  The text under Technical Merits heading of the \<description\> tag should be parsed into the \<advantage\> tag.

- Emory University
  \<inventor-first-name\> tag contains both the first name and last name. The last name should be parsed into the \<inventor-last-name\> tag. The Highlights section under \<summary\> belongs in the \<advantage\> tag.

- Harvard University
  The \<summary\> tag contains an INNOVATIONS & ADVANTAGES section heading which should be parsed into the \<advantage\> tag.

- Medical University of South Carolina
  The Advantages section under the \<summary\> tag can be put into the \<advantage\> tag.

- University of Cincinnati
  In the \<summary\> tag, if there exists an Advantages heading, the text under it should be put into the \<advantage\> tag. Some cases do not have this section.

- University of Michigan
  The \<description\> tag also contains and Advantages heading. The text under that heading should be move into the \<advantage\> tag.

- Vanderbilt University
  In the \<summary\> tag, there are links to the Inventors with their full names. It seems there can be more than one inventor. Although the links appear to be broken, the inventors' names can be parsed into the \<inventor-first-name\> and \<inventor-last-name\> tags.

Inventor

Implemented Features:

- Created author node for inventors
- Parsed inventors' names from the \<description\> and \<summary\> tags for the following universities: Columbia, Tufts, Vanderbilt
  Warning: The parse string may contain multiple inventors and even \<a href\> tags
- Parsed out Inventor's first, middle, and last names from cases where \<inventor-first-name\> contain the full name and \<inventor-last-name\> is empty
- Formatted the inventor name to display as Last, First Middle

Unimplemented Features:

- Parse out name from \<a href\> Name \</a\> tags
- Create separate person (author) nodes for inventor strings with multiple inventors
- Smush/collapse inventors

Implementation Detail: Case Analysis

Testing Data: Inventors.xml

Advantages

Unimplemented Features:

- Parse advantages by University