Mindsumo Challenge: Create a Data Visualization to Navigate Educational Standards

Data Samples Overview
https://www.mindsumo.com/contests/educational-data-visualization

***Abstract*—This overview provides information on how to interpret the public data samples that have been provided for this Mindsumo challenge. This includes information on file formats and relevant fields. All of this data is publicly available.**

CONTENTS

I. INTRODUCTION

For this challenge you will be working with multiple collections of education standards that were designed for a particular subject, e.g. Math, English, Science. These standards collections are hierarchical and each may have their own arbitrary number of levels. For example, the Common Core State Standards (CCSS) for Math[1] are arranged in a structure as follows: Subject (Math) $\rightarrow$ Domains $\rightarrow$ Clusters $\rightarrow$ Standards. Part of the challenge then is to provide efficient ways for viewers to access and read the collection's content data, i.e. labels or sentences that represent a domain, cluster or standard, e.g.

- Domain entry ($n_{1.1}$): "Number and Operations in Base Ten"
- Cluster entry ($n_{1.1.1}$): "Use place value understanding and properties of operations to perform multi-digit arithmetic"
- Standard entry ($n_{1.1.1.1}$): "Use place value understanding to round whole numbers to the nearest 10 or 100"

Another part is being able to navigate and interact with the collection's structural, hierarchy data (i.e. which elements are parents or children of each other).

- (parent,child): ($n_{1.1}$, $n_{1.1.1}$)
- (parent, child): ($n_{1.1.1}$, $n_{1.1.1.1}$)

[1]http://www.corestandards.org/Math/

The challenge is also to provide efficient visualization of standards alignment or "crosswalks" that have been provided between multiple collections. Standards alignment is a process where standards describing similar concepts across multiple standards collections are correlated into a crosswalk. The scenario we are most interested in is one where we have a source collection (S) and multiple target collections (T). We would like to see effective visualizations that can allow a user to have both the source and the multiple targets visible, along with the crosswalks so that they can explore the data visually. Crosswalks define collections of alignment links between source and target entries, e.g.

- (source,target): ($n_{1.1.1}$, $n_{2.1.1.1}$)

II. DATA SAMPLE SETS

Provided are education standards that have been publicly published on the Achievement Standards Network (ASN)[2]. This includes both the standards content/hierarchical structure as well as specified crosswalks/alignments.

A. Sample 1: Mathematics

The first data sample set is taken from the Achievement Standards Network (ASN) public crosswalk data for Math collections. The scenario is a hierarchical source collection S_1 that is aligned to multiple target hierarchical collections $\{T_1, T_2\}$. The following table defines the details for each element of this sample.

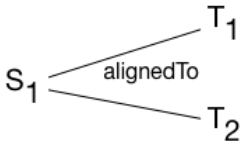

Fig. 1. Math Source, Target Alignment Sample

Id	Subject	Data File	Description
S_1	Math	s1.json	Common Core State Standards (CCSS) for Math
T_1	Math	t1.json	American Association of School Librarians Standards (AASL) for the 21st Century Learner
T_2	Math	t2.json	South Carolina State Standards (SCSS)
$T_1 - S_1$	Math	t1-s1.csv	AASL - CCSS
$T_2 - S_1$	Math	t2-s1.csv	SCSS - CCSS

[2]http://www.achievementstandards.org

B. Sample 2: English Language Arts (ELA)

The second data sample set is taken from the Achievement Standards Network (ASN) public crosswalk data for English Language Arts (ELA) collections. The scenario is a hierarchical source collection S_2 that is aligned to multiple target hierarchical collections $\{T_3, T_4\}$. The following table defines the details for each element of this sample.

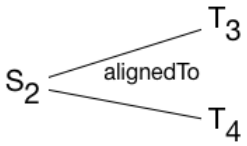

Fig. 2. Math Source, Target Alignment Sample

Id	Subject	Data File	Description
S_2	ELA	s2.json	Common Core State Standards (CCSS) for English Language Arts and Literacy in History/Social Studies, Science, and Technical Subjects
T_3	ELA	t3.json	Michigan State Standards (MSS)
T_4	ELA	t4.json	South Carolina State Standards (SCSS)
$T_3 - S_2$	ELA	t3-s2.csv	MSS - CCSS
$T_4 - S_2$	ELA	t4-s2.csv	SCSS - CCSS

III. Data Format and Fields

In this section we discuss the two file formats that communicate:

1) The collection element content data and hierarchical structural data
2) The crosswalk/alignment structural data

More detailed information on these formats can be found at the ASN machine readable source papers [1].

A. Content/Structural Data Format

The content/structural data files {s1.json, s2.json, t1.json, t2.json, t3.json, t4.json} as the extension suggests are all JavaScript Object Notation or JSON[3] files. The key values were selected by the ASN from open data formats. There are two types of entries in this file:

1) Standards Document
2) Standards Statement

The first JSON entry is a “Standards Document" element which provides meta data information about the collection, including a source URL where the collection can be publicly found. For this exercise, we can skip this entry. The second entry is also a “Standards Document" entry and represents the root node of the collection.

The content data for this root node can be found in the title and description fields.

[3]https://en.wikipedia.org/wiki/JSON

Listing 1. Root node content data

```
"http://asn.jesandco.org/resources/D10003FB" : {
  ...
  "http://purl.org/dc/elements/1.1/title" : [
     { "value" : "Common Core State Standards for Mathematics", "type" : "literal",
           "lang" : "en-US" }
    ],
  "http://purl.org/dc/terms/description" : [
     { "value" : "These Standards define what students should understand...",
           "type" : "literal", "lang" : "en-US" }
    ],
  ...
},
```

The hierarchy structural data for this node is expressed using the Gateway to Educational Materials (GEM)[4] predicate ’hasChild’.

Listing 2. Hierarchy structural data

```
"http://asn.jesandco.org/resources/D10003FB" : {
  ...
  "http://purl.org/gem/qualifiers/hasChild" : [
     { "value" : "http://asn.jesandco.org/resources/S2366905", "type" : "uri" },
     { "value" : "http://asn.jesandco.org/resources/S1143409", "type" : "uri" },
     { "value" : "http://asn.jesandco.org/resources/S114340A", "type" : "uri" },
     ...
    ]
    ...
}
```

This essentially links within the document, defining the (parent,child): ($n_{1.1}$, $n_{1.1.1}$) relationship. Note that the identifying value used is the URI key to the JSON object, e.g. “http://asn.jesandco.org/resources/S2366905".

Standards statement JSON entries have content fields for statement URI (identifier for the object entry), statement notation (short code), description and hierarchy structural data as well, for example:

Listing 3. Hierarchy structural data

```
"http://asn.jesandco.org/resources/S114340E" : {
    ...
    "http://purl.org/ASN/schema/core/statementNotation" : [
      { "value" : "CCSS.Math.Content.K.CC.A", "type" : "literal" }
    ],
    ...
    "http://purl.org/dc/terms/description" : [
      { "value" : "Know number names and the count sequence.", "type" : "literal",
            "lang" : "en-US" }
    ],
    ...
    "http://purl.org/gem/qualifiers/hasChild" : [
      { "value" : "http://asn.jesandco.org/resources/S1143417", "type" : "uri" },
      { "value" : "http://asn.jesandco.org/resources/S1143418", "type" : "uri" },
      { "value" : "http://asn.jesandco.org/resources/S1143419", "type" : "uri" }
    ]
    ...
  },
}
```

B. Crosswalk/Alignment Data Format

The content/structural data files {t1-s1.csv, t2-s1.csv, t3-s2.csv, t4-s2.csv} as the extension suggests are all Comma separated value or CSV[5] files. The header entries were defined by the ASN community. They include:

- subjectURI
- subjectLabel
- subjectNotation
- subjectGrade
- predicateCURIE
- predicateLabel
- objectURI
- objectLabel

[4]http://www7.scu.edu.au/1897/com1897.htm

[5]https://en.wikipedia.org/wiki/Comma-separated_values

- objectNotation
- objectGrade
- relationshipID

Each row is a pointer from a node in one collection to a node in the other collection. This can occur from any source level to any target level. The two critical structural fields here are subjectURI and objectURI. For example, this entry aligns a T_1 node to a S_1 node.

Listing 4. Crosswalk data

```
subjectURI, ... objectURI, ...
http://asn.jesandco.org/resources/S113BA38, ...,
      http://asn.jesandco.org/resources/S11434E7, ...
```

In order to integrate a navigational approach to both the content/hierarchical and crosswalk data for multiple source and target collections, both formats need to be parsed and represented. Challenge entries may choose to unify these source datasets into a new, single source JSON representation. Additionally, challenge entries may choose to index content (e.g. description fields) to implement free text searching or to leverage additional fields for handy filtering, e.g. education level, statement notation codes.

REFERENCES

[1] S. A. Sutton and D. Golder, "Achievement standards network (ASN): an application profile for mapping k-12 educational resources to achievement standards," in *International Conference on Dublin Core and Metadata Applications*, 2008, pp. 69–79.