A XPath Debugger based on Fuzzy Chance Degrees

Jesús M. Almendros-Jiménez
Dept. of Languages and Computation
University of Almeria, Spain
Email: jalmen@ual.es

Alejandro Luna Tedesqui
Dept. of Computing Systems
University of Castilla-La Mancha, Spain
Emails: Alejandro.Luna@alu.uclm.es

Ginés Moreno Valverde
Dept. of Computing Systems
University of Castilla-La Mancha, Spain
Emails: Gines.Moreno@uclm.es

"In this Poster we present a method for debugging XPath. We will describe how XPath expressions can be manipulated in order to obtain a set of alternative XPath expressions that match to a given XML document. For each alternative XPath expression we will give a change degree that represents the degree in which the expression deviates from the initial expression. Thus, our work is focused on providing to the programmer a repertoire of paths that (s)he can use to retrieve answers. The approach has been implemented and tested."

Our debugging method acts on a given initial XPath query $\boldsymbol{Q}$ preceded by the [**DEBUG = r**] command, where $\boldsymbol{r}$ is a real number in the unit interval used at debugging time for labeling deletion and jumping actions. So, assume that we plan to process "**\[DEBUG=0.5]/bib/book/title**" w.r.t. the following XML document:

```
<bib>
   <name>Classic Literature</name>
   <book year="2001" price="45.95">
      <title>Don Quijote de la Mancha</title>
      <author>Miguel de Cervantes Saavedra</author>
      <references>
         <novel year="1997" price="35.99">
            <name>La Galatea</name>
            <author>Miguel de Cervantes Saavedra</author>
            <references>
               <book year="1994" price="25.99">
                  <title>Los trabajos de Persiles y Sigismunda</title>
                  <author>Miguel de Cervantes Saavedra</author>
               </book>
            </references>
         </novel>
      </references>
   </book>
   <novel year="1999" price="25.65">
      <title>La Celestina</title>
      <author>Fernando de Rojas</author>
   </novel>
</bib>
```

Our technique produces a set of alternative queries $\boldsymbol{Q_1, \ldots, Q_n}$ (each one adorned with attributes and special keywords informing about all changes that deviates $\boldsymbol{Q_i}$ from $\boldsymbol{Q}$) packed into an output XML document like the following one, where the set of proposals is sorted w.r.t. a CD key meaning that, as much changes are performed on $\boldsymbol{Q_i}$ and as more *traumatic* they are w.r.t to $\boldsymbol{Q}$, then its *"Chance Degree"* becomes lower according a policy based on the *product fuzzy logic*:

```
<result>
   <query cd="1.0">/bib/book/title</query>
   <query cd="0.8" book="novel">/bib/[SWAP=0.8]novel/title</query>
   <query cd="0.5" book="//">/bib/[JUMP=0.5]//title</query>
   <query cd="0.5" bib="//">/[JUMP=0.5]//book/title</query>
   <query cd="0.45" book="" title="name">/bib/[DELETE=0.5][SWAP=0.9]name</query>
   <query cd="0.225" bib="" book="//" title="name">/[DELETE=0.5][JUMP=0.5]//[SWAP
       =0.9]name</query> ...
</result>
```

Swapping case: Instead of $\boldsymbol{tag_i}$ we find $\boldsymbol{tag'_i}$ at level $\boldsymbol{i}$ in the input XML document $\boldsymbol{D}$, being $\boldsymbol{tag_i}$ and $\boldsymbol{tag'_i}$ two similar terms with similarity degree $\boldsymbol{s}$. Then, we generate an alternative query by adding the attribute $\boldsymbol{tag_i}$=$\boldsymbol{tag'_i}$ and replacing in the original path the occurrence "$\boldsymbol{tag_i}$" by "$\boldsymbol{[SWAP=s]tag'_i}$".

The second query proposed this case:

<query cd="0.8" book="novel"»/bib/[SWAP=0.8]novel/title</query>

Let us observe that : **1)** we have included the attribute **<<book="novel">>** in order to suggest that instead of looking now for a book, finding a novel should be also a good alternative, **2)** in the path we have replaced the tag *book* by *novel* and we have appropriately annotated the exact place where the change has been performed with the annotation [**SWAP=0.8**] and **3)** the *cd* of the new query has been adjusted with the *similarity* degree 0.8 of the exchanged tags.

Execution of query "/bib/[SWAP=0.8]novel/title"

```
<result>
   <title rsv="0.8">La Celestina</title>
</result>
```

Jumping case: Even when $\boldsymbol{tag_i}$ is not found at level i in the input XML document D, $\boldsymbol{tag_{i+1}}$ appears at a deeper level (i.e, greater than i) in a branch of D. Then, we generate an alternative query by adding the attribute tag_i="//", which means that $\boldsymbol{tag_i}$ has been jumped, and replacing in the path the occurrence **"$\boldsymbol{tag_i}$"** by **"[JUMP=r]//"**, being r the value associated to DEBUG.

Execution of query "/bib/[JUMP=0.5]//title"

```
<result>
   <title rsv="0.5">Don Quijote de la Mancha</title>
   <title rsv="0.5">La Celestina</title>
   <title rsv="0.03125">Los trabajos de Persiles y Sigismunda</title>
</result>
```

Deletion case: This situation emerges when at level i in the input XML document D, we found $\boldsymbol{tag_{i+1}}$ instead of $\boldsymbol{tag_i}$. So, the intuition tell us that $\boldsymbol{tag_i}$ should be removed from the original query $\boldsymbol{Q}$ and hence, we generate an alternative query by adding the attribute $\boldsymbol{tag_i}$**=""** and replacing in the path the occurrence "$\boldsymbol{tag_i}$" by **"[DELETE=r]"**, being **r** the value associated to DEBUG.

This situation is illustrated by the query **"/bib/[DELETE=0.5][SWAP=0.9]name"**, where the deletion of the tag book is followed by a swapping of similar tags *title* and *name*. The **cd 0.45** associated to this query is defined as the product of the values associated to both **DELETE** (0.5) and **SWAP** (0.9), and hence the chance degree of the original one is lower than the previous examples. The execution of our new query is able to retrieve the information contained in the first branch of the input XML document.

Execution of query "/bib/[DELETE=0.5][SWAP=0.9]name"

```
<result>
   <name rsv="0.45">Classic Literature</name>
</result>
```

As we have seen in the previous example, the combined use of one or more debugging commands (**SWAP**, **JUMP** and **DELETE**) is not only allowed but also frequent. In other words, it is possible to find several debugging points. We can see the execution of the query "/[DELETE=0.5][JUMP=0.5]//[SWAP=0.9]name", with *cd* 0.225 is quite low, and therefore the change degree is low, since it has been obtained by multiplying the three values associated to the deletion of the tag *bib* (0.5), jumping the tag *book* (0.5) and the swapping of *title* by *name* (0.9).

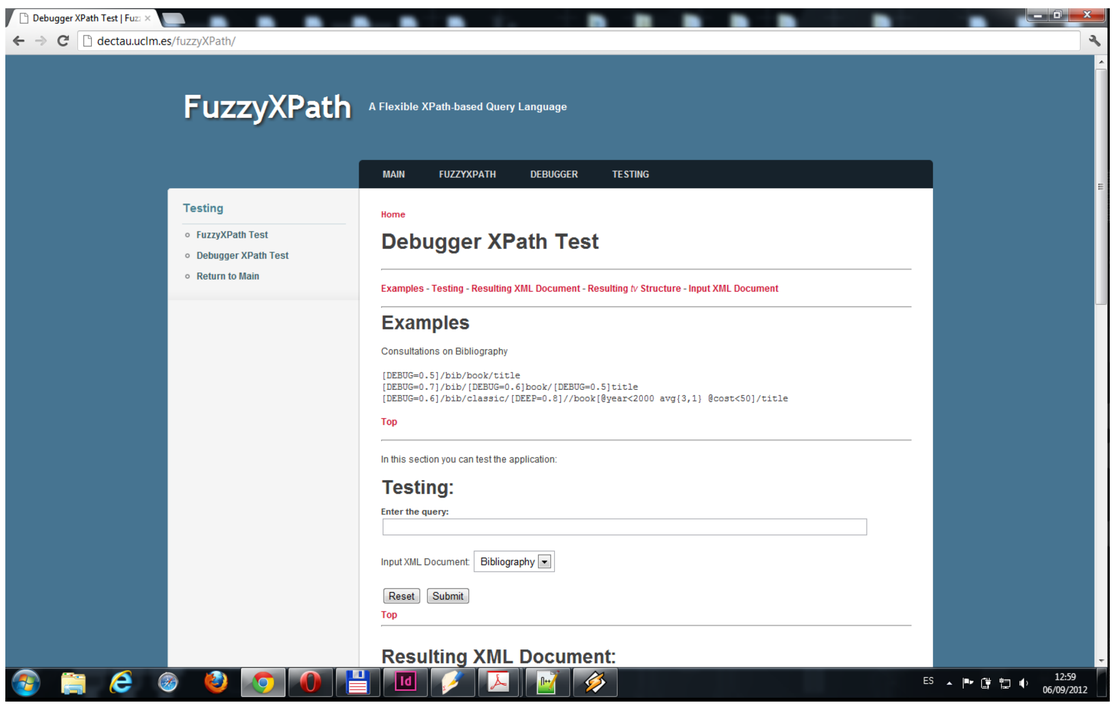
Execution of query "/[DELETE=0.5][JUMP=0.5]//[SWAP=0.9]name"

```
<result>
   <name rsv="0.225">Classic Literature</name>
   <name rsv="0.028125">La Galatea</name>
</result>
```

So, after executing all these queries, please, consult and visit:

http://dectau.uclm.es/fuzzyXPath/

where it is possible too to perform an on-line debugging session.



CONCLUSIONS.

The result of the debugging process of a XPath expression is a set of alternative queries, each one associated to a chance degree. We have proposed **JUMP**, **DELETE** and **SWAP** operators that cover the main cases of programming errors when describing a path about an XML document. We have implemented and tested the approach. The approach has a fuzzy taste in the sense that XPath expressions are debugged by relaxing the path expression assigning a change degree to debugging points. The fuzzy taste is also illustrated when executing the annotated XPath expressions in a fuzzy based implementation of XPath.

In the near future, we plan to implement our approach in existent XPath/XQuery implementations. We are now working in the implementation of fuzzy XPath with XQuery. The idea is to define XQuery functions to represent fuzzy operators: DEEP, DOWN, JUMP, DELETE and SWAP, avg, etc, in order to execute fuzzy XPath with any XQuery interpreter.

Acknowledgements: Work partially supported by the EU, under FEDER, and the Spanish Science and Innovation Ministry (MICINN) under grant TIN2008-06622-C03-03, as well as by Ingenieros Alborada IDI under grant TRA2009-0309, and the JUNTA ANDALUCIA administration under grant TIC-6114.