

A Probabilistic XML Merging Tool



Talel Abdessalem
Télécom ParisTech
Paris, France

Mouhamadou Lamine BA
Université Cheikh Anta DIOP
Dakar, Senegal

Pierre Senellart
Télécom ParisTech
Paris, France



<http://dbweb.enst.fr/>

What this tool aim at...

- Representing the outcome of semi-structured documents integration as a probabilistic tree
- Evaluating the uncertainty (modeled as probability values) of the result of the merge
- Querying the probabilistic repository with a subset of the XPath query language

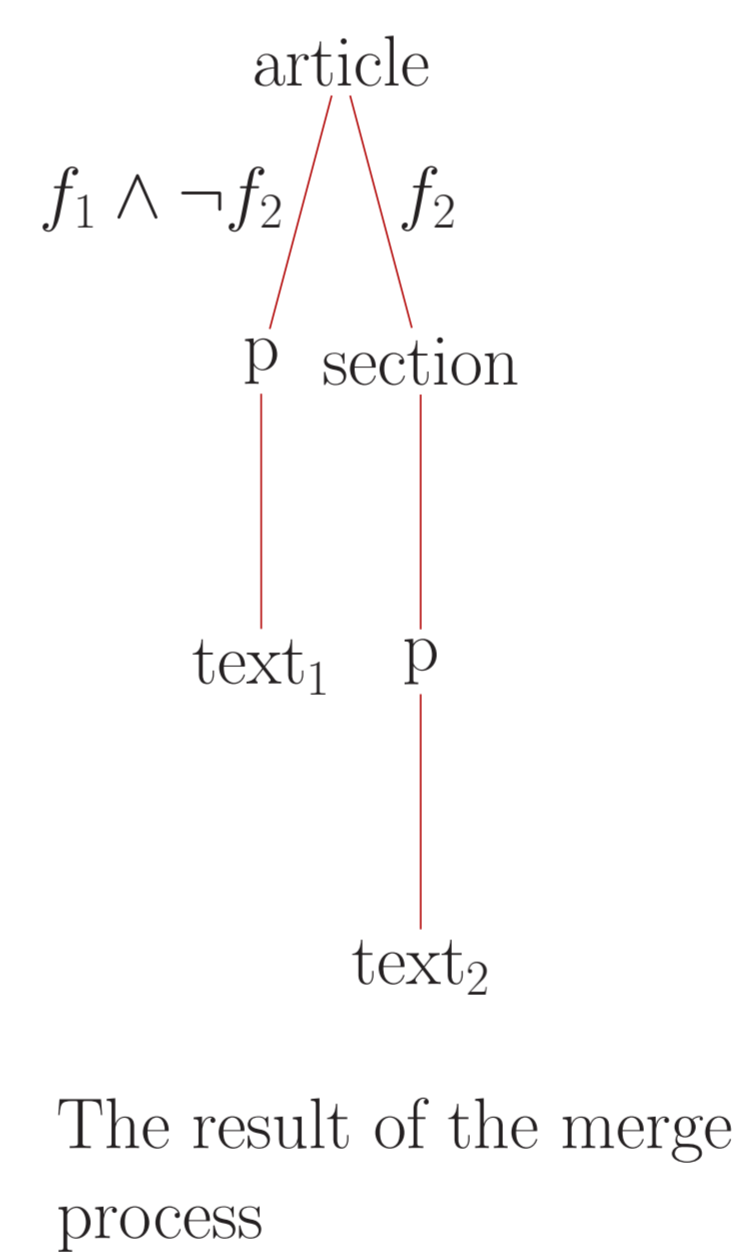
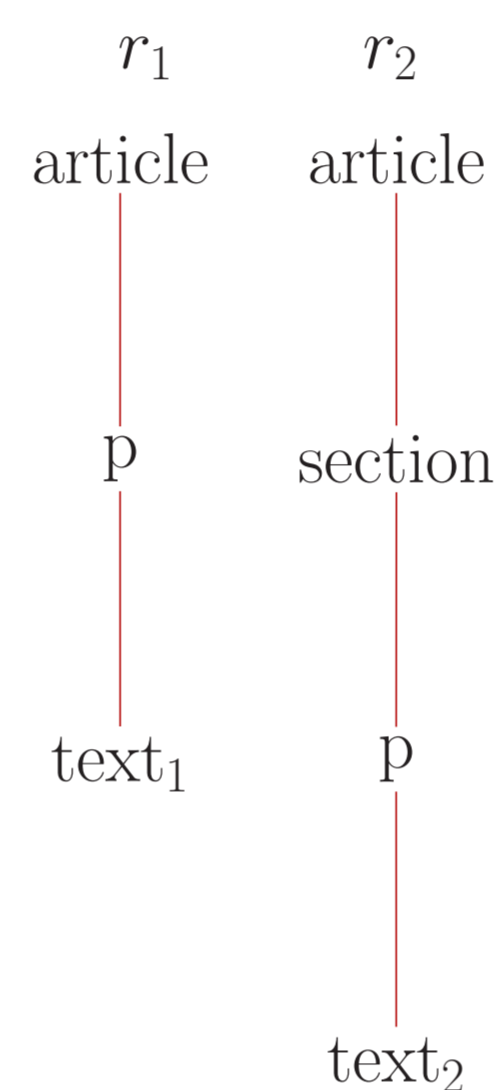
Application domain: [Wikipedia revisions](#)

The tool enables merging the revisions of a given Wikipedia page with:

- an efficient evaluation of the uncertainty of the obtained result
- an automatic management of conflicts.

Merging of Wikipedia revisions

- A two-way tree merging technique for P-Documents
- Two steps: [Matching of Revisions](#) and [Merging Matches](#)



1. Matching of Revisions

Input: two revisions r_{k-1} and r_k and their associated event formula.

Output:

- Deleted nodes x : $x \in r_{k-1}$ and x has no match in r_k .
- Added nodes x : $x \in r_k$ and x has no match in r_{k-1} .
- Matched couples (x, y) : $x \in r_{k-1}$ and $y \in r_k$ match.

2. Merging Matches

- Deleted nodes:

$$fie_{new}(x) = fie_{old}(x) \wedge (\neg f_k)$$

- Matched couple:

$$fie_{new}(x) = fie_{old}(x)$$

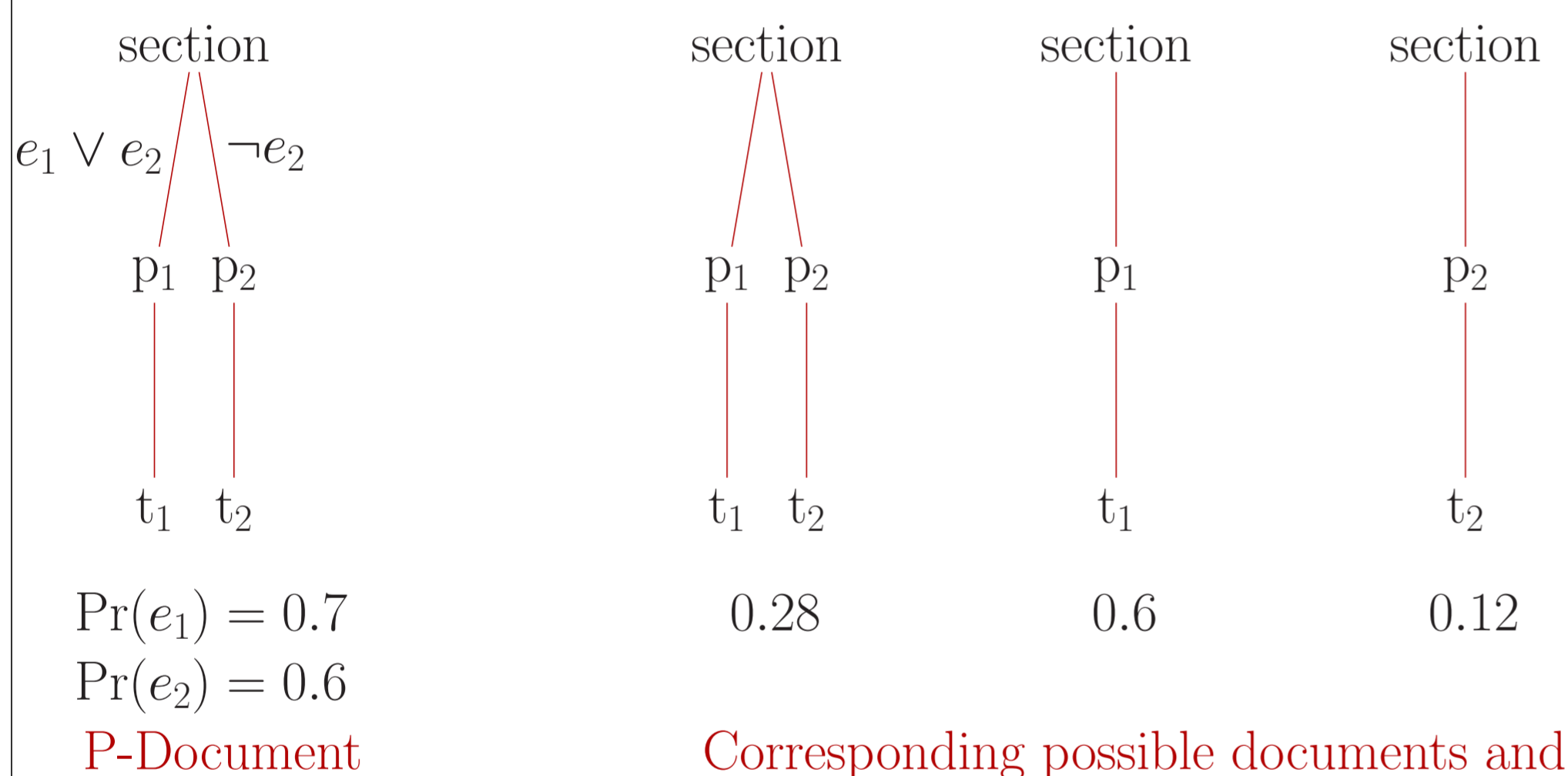
- For added nodes:

$$fie_{new}(x) = f_k$$

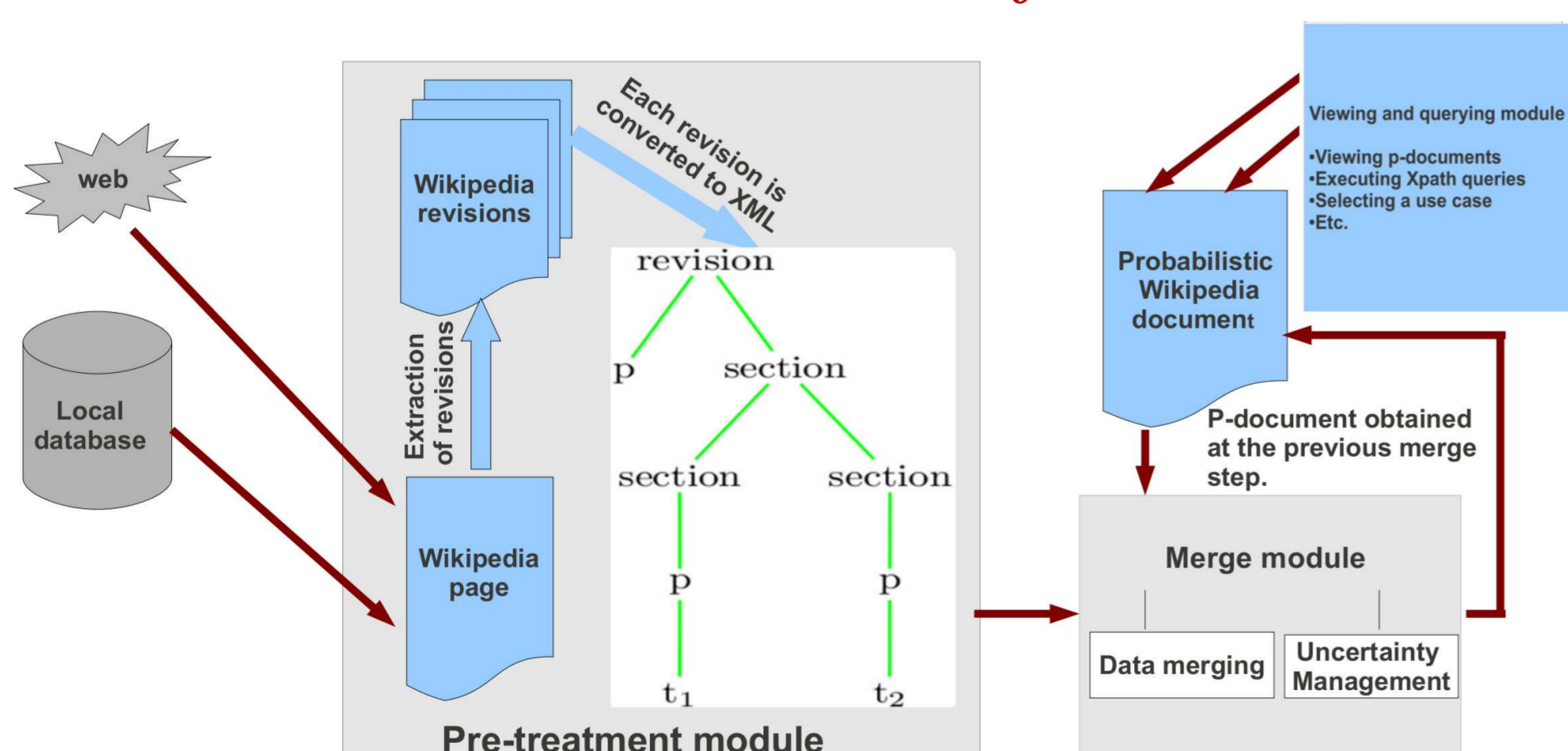
or

$$fie_{new}(x) = fie_{old}(x) \vee f_k$$

Probabilistic Documents



Architecture of the system



Description of the system

- [System for managing Wikipedia documents.](#)

Features

- A keyword-based search engine for Wikipedia pages
- Extracting the revisions of a given page
- Selecting the list of revisions to merge
- Building one's own Wikipedia article
- Displaying the result of the merge
- Demonstrating a certain number of use cases
- Using a subset of XPath query language