A Framework for Usage-based Document Reengineering

Madjid SADALLAH - *CERIST, Algeria*	Benoît ENCELLE - *University of Lyon, France*
Azze-Eddine MAREDJ - *CERIST, Algeria*	Yannick PRIÉ - *University of Nantes, France*

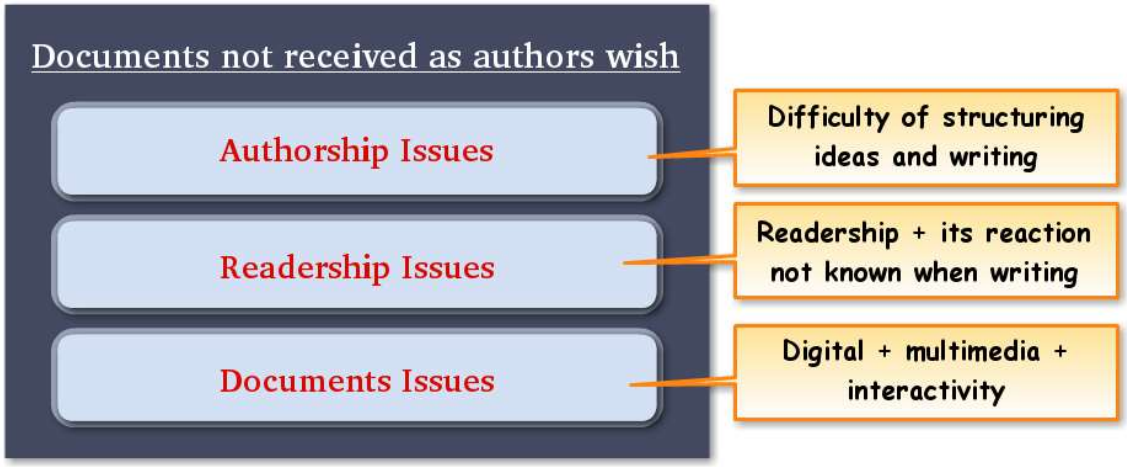
Outline

1. Context / Motivations
2. Usage-based Document Reengineering
3. Framework for Usage-based Document Reengineering
4. Conclusion and Future Work

1

Context / Motivations

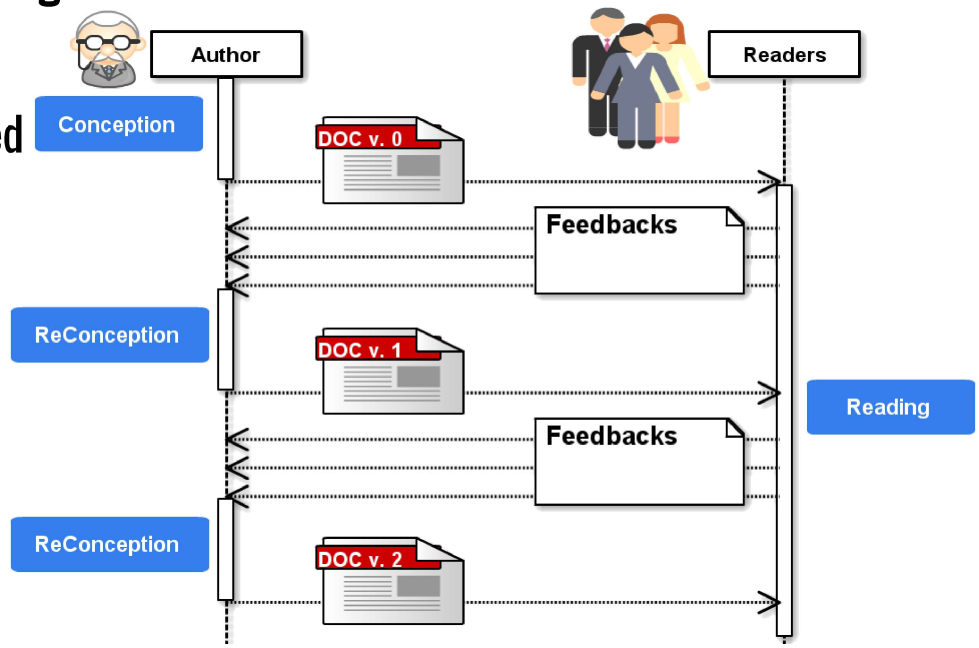
- General context : document engineering –> digital publishing and reading.
- Technical context : LMS **Claire** (Community Learning through Adaptive and Interactive multichannel Resources for Education. http://www.projet-claire.fr). Towards a simple, yet robust tool for authoring, improving and disseminating educational content.
- Assist authors to better convey knowledge contained within their docs.
- Difficulty to design documents that are received the way the author wishes. Why?
- Writing issue? reading issue? document nature issue? ALL of these?



2

Usage-based Document Reengineering

- Digital facilities: a persistent, two-ways communication established between authors and readers
- Idea : two-ways communication → enable author to consider reader usages feedbacks as a knowledge source for his documents reconception.
- Document reconception = document structures and content revision (reengineering).
- That's we call **Usage-based Document Reengineering**.
- We aim to be able to:
  - Offer means to enable such a usage-based document reengineering.
  - Assist authors in performing the reengineering tasks.



Reengineering

"Reengineering, also known as both renovation and reclamation, is the examination and alteration of a subject system to reconstitute it in a new form and the subsequent implementation of the new form" (Chikofsky et al., 1990).

Document reengineering

- Reengineering applied on documents *structures* and *content*.
- Possible: digital documents are nowadays no longer single-version with static content, they are "live": multimedia, multiuser, dynamic and thus multi-version (Balinsky, 2011).
- Our focus: facilitate document appropriations by readers.

Usage-based document reengineering (digital publishing and reading context)

A document reengineering performed in response to readers' usages feedback:

- explicit usages : **annotations**
- implicit usages : **reading traces**

Readers' Usages

Annotation

Any information provided by a user that is associated with a whole document or a part of it.

Reading Trace

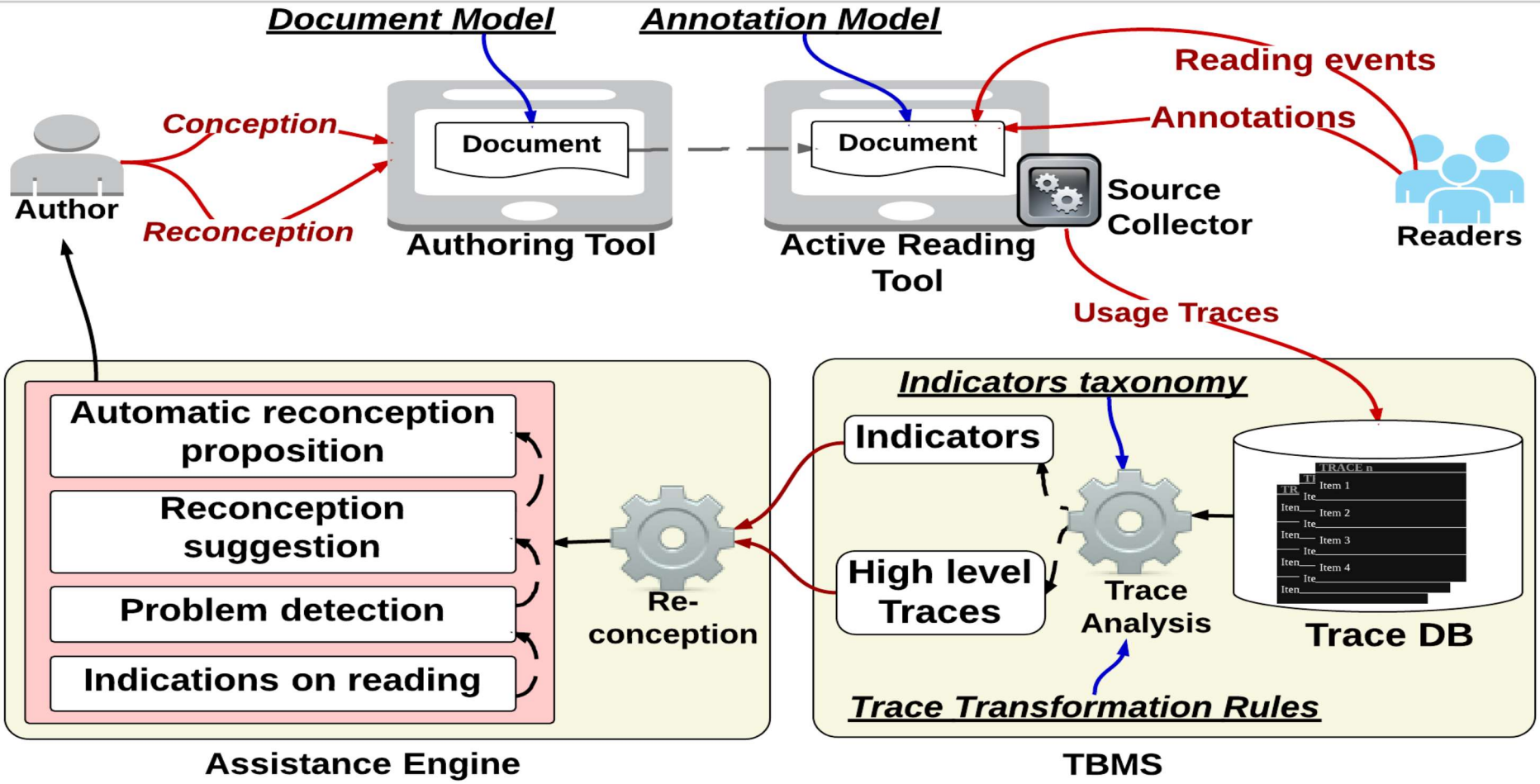
Temporal sequence of observed elements recorded from interactions between a reader and a document, through a reading tool.

Requirements for usage-based reengineering

- Authoring tool to conceive/reconceive documents
- Reading tool that enables active reading + captures readers usages
- Reengineering facilities to interpret usages and generate/perform reconception.

3

Framework for Usage-based document reengineering

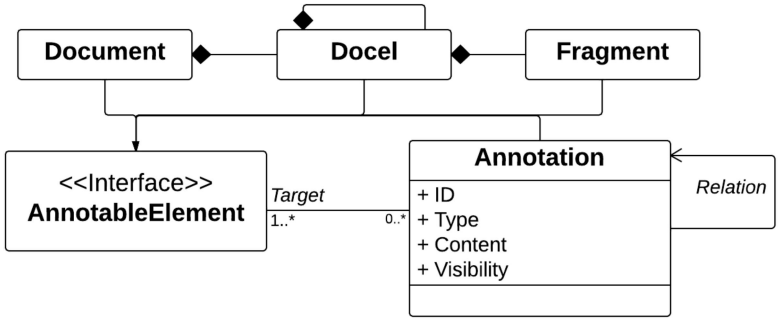


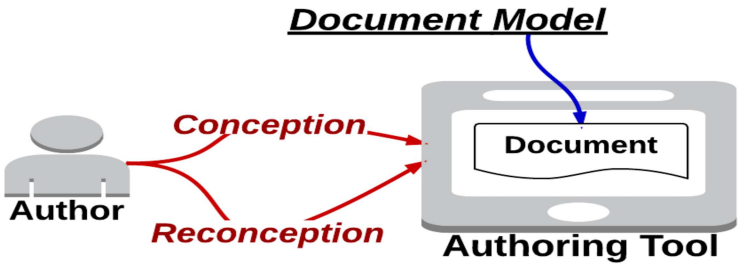
Overview of the technical framework

Auhtoring tool

Means to conceive and reconceive documents.

Document



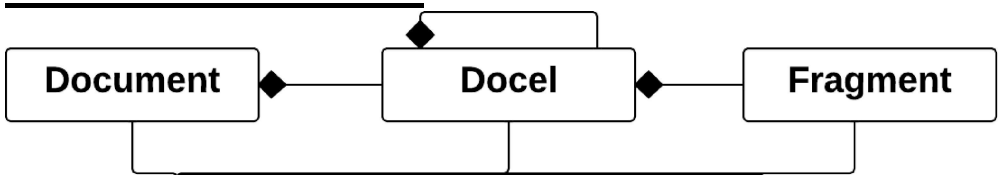


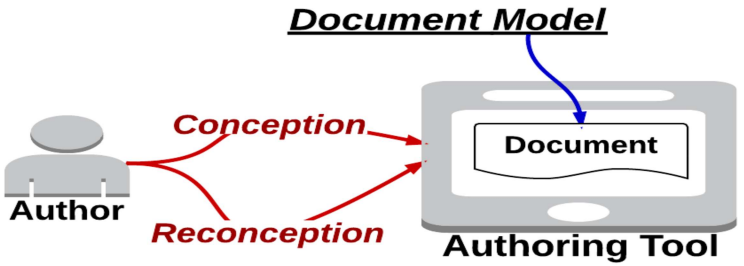
Auhtoring tool

Means to conceive and reconceive documents.

Document

- Reconception: document content + structures.
- Generic document model





- **Document** composed of docels.
- **Docel** = building blocks. Have different types of attributes (composition, placement, synchronization and behavior).
- **Fragment** = a logical part of a docel. Defined using spatio-temporal coordinates.

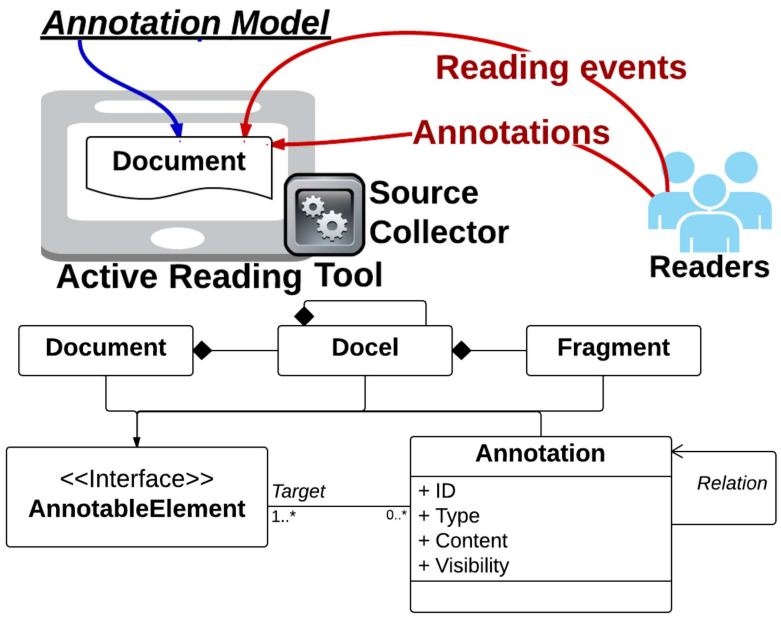
Reading tool

Active-reading capabilities.

Source Collector : collects the obsels (*primary trace*) resulted from the document active reading.

Annotation

- An annotation target: **annotable element**–> one or more document elements and/or fragments and/or annotations.
- Each annotation has one and only one type. Within **Claire** : *Question, Form error, Content error, Comment, I understood, I did not understand, Lecture notes* (personal notes) and *Other* + *Highlighting* + *Linked annotation* (annotate an annotation).
- Visibility: control annotation availability to users. Within **Claire** : *private, to author/reviewer, group* and *public*.



TBMS: Trace-Based Management System

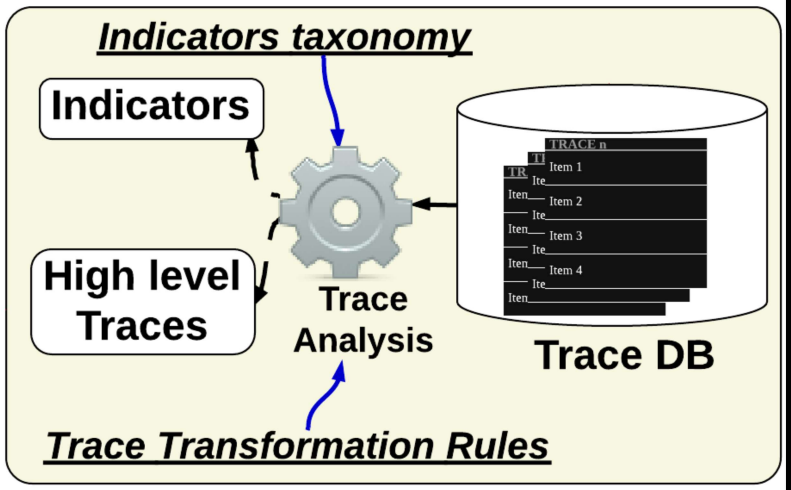
Store and process the collected data (primary traces).
Processing results: **Indicators** and **High level traces**.

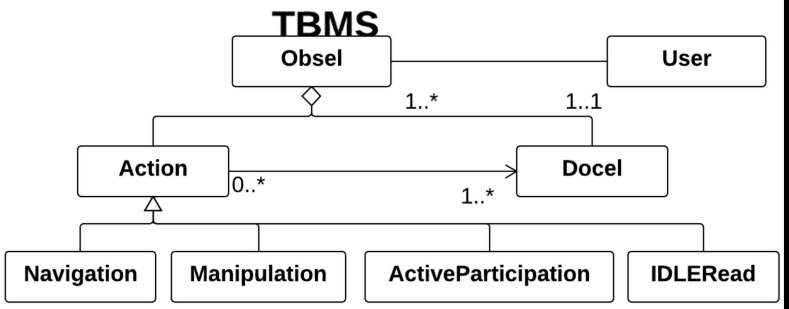
Trace model

- Based on modeled Traces (see SILEX, LIRIS, CNRS).
- **Obsel** : *observed element* associated to a user, connects a specific action with a document element.

Obsel types

- **Navigation.** following links, visiting specific URLS, scrolling (spatially and/or shifting in time) and moving back and forward in navigation history.
- **Manipulation.** On document content (e.g. select, find, print, zoom, copy and bookmark + media controls: play/pause/stop, seek, etc.) and context (e.g. activating system interface to open/close/download the document).





TBMS: Trace-Based Management System

- **Active participation.** adding/altering/deleting one's annotations, annotating/opening/closing an annotation, highlighting, etc.
- **IDLERead.** Passive reading, no interaction, maybe inactivity or absence.

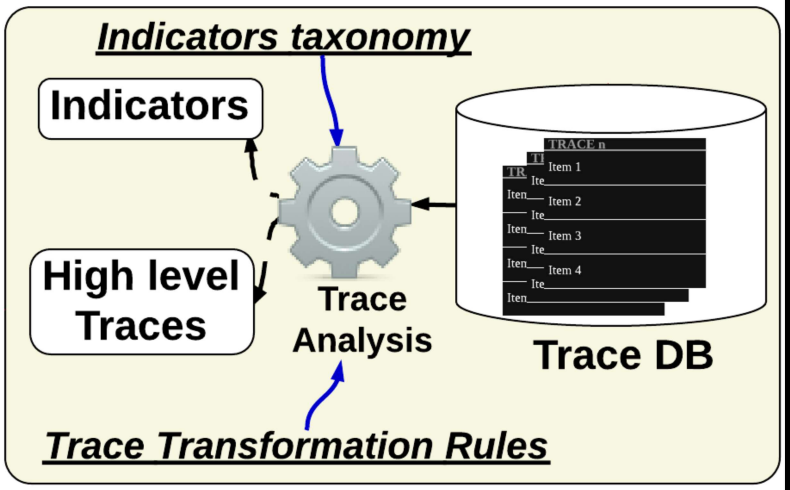
Computation on primary traces results

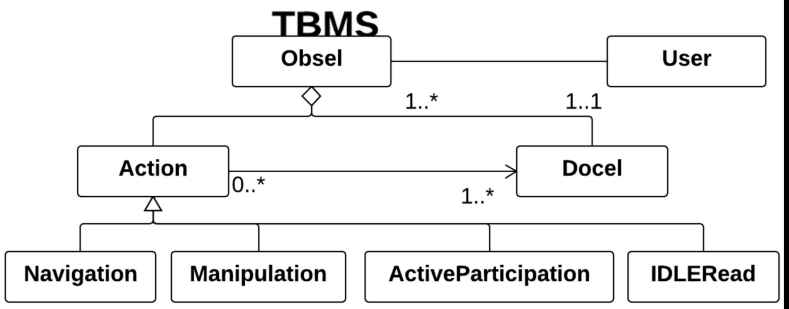
High level traces

The result of a *transformation process* performed on the primary trace to interpret and abstract it. Ex: *filtering*, *rewriting* and *aggregating* obsels.

Indicators

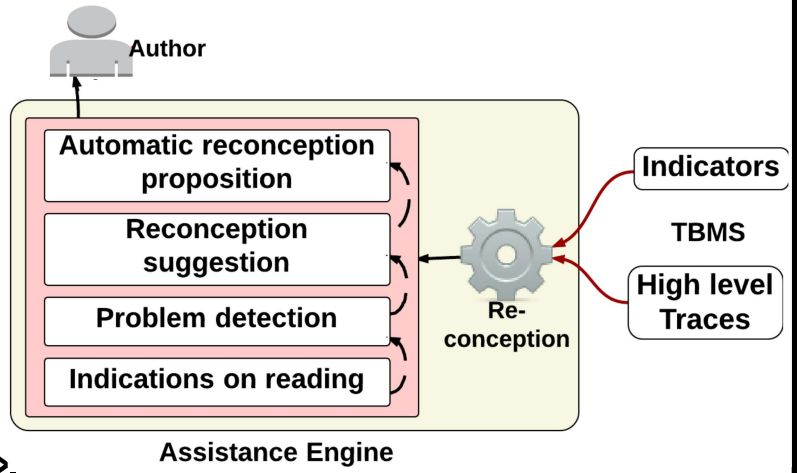
Variables computed to characterize readers' interaction against a specific monitored feature or event. Ex: unread sections, (un)visited links, spent time.





Assistance engine

Analyses data provided by the TBMS to assess possible and appropriate document reconceptions.
Arbitrary set of feedbacks (from a single reader, a given group of readers or the entire readership). End result: a new version of the document possibly subject to further revisions.



4 levels of assistance. Level <**N**> uses data provided by level <**N - 1**>.

- *Level 0* : **Indications on reading** . AE computes and presents the author with indications on how the document has been read.
- *Level 1* : **Problem detection** . AE detect problems in the reading process but not give any suggestion on how to fix them.
- *Level 2* : **Reconception suggestion** . AE detects problems, supply suggestions on fixing them nut is unable by itself to achieve them.
- *Level 3* : **Automatic reconception proposition** . AE detecta and resolves problems automatically. The reconception can be reviewed and validated by the author.

4

Conclusion & Future Work

2 issues

- How to reconceive documents by exploiting readers' feedbacks
- How to assist authors to achieve such reconceptions.

Proposal

A framework for document reengineering that uses readers' usage feedbacks (reading traces and annotations) and offers authors various levels of assistance.

Future Work

Conception of suitable means and tools to assess reconceptions, using the primary traces and going through the suitable trace transformations and indicators computation.

Next within Claire

- Interviews with authors to identify the actual reconception needs.
- Precise/specialize the different associated models.
- A meaningful set of transformations and indicators for enhancing documents.

THANK YOU!