Université Abderrahmane Mira - Béjaïa
Faculté des Sciences Exactes
Département d'Informatique

Soutenance de Thèse de Doctorat en Science
Spécialité: Informatique
Option: Cloud Computing

Models and Tools for Usage based e-Learning Documents Reengineering

Présentée par: **Madjid SADALLAH**

Directeur de thèse: Professeur **Yannick PRIE** (*Université de Nantes*)

25 Avril 2019

Roadmap

Background & Motivation

- With digitization and ICTs
  - Popularization of e-learning platforms
  - Learning: distant, online and self-directed
  - Positive implications on teaching & learning
- Important responsibility on leaners. But studies found
  - Learners' comprehension issues (e.g. disorientation and cognitive load)
  - For extracting a correct and complete knowledge from e-courses:

Learners need support → From instructors/course authors

Background & Motivation

- Comprehension: greatly depends on course quality
- To ensure and maintain quality
  - Frequent revisions
  - According to learners' needs
- Authors make infrequent and superficial revisions:
  - Learners' comprehension problems/barriers?
  - Challenging parts/aspects?
  - Right corrective measures?

Authors also need to be assisted

Background & Motivation

- No efficient assistive tools
  - readability formulas
  - human-oriented methods (exams, questionnaires)
- We propose an usage-based assistance
  - Learning platforms are shifted with logging capabilities
  - Learners' traces as a source of knowledge
  - To not require analytics skills:

Automated analytics tools with assistive capabilities

Research aim

Investigate the use of analytics on learners' traces to

- study learners' **reading behavior**
- detect their **comprehension issues**,
- Identify course **parts and aspect** related to those issues,
- **suggest remediation actions** to authors to improve their contents

1. Increase in data captured ('big data')
2. Advancement in computing facilities
3. Popularization of analysis methods

Learning analytics

Measurement, *collection*, *analysis* and *reporting* of data about learners and their contexts, for purposes of understanding and optimizing learning and its environments (Siemens 2013)

Roadmap

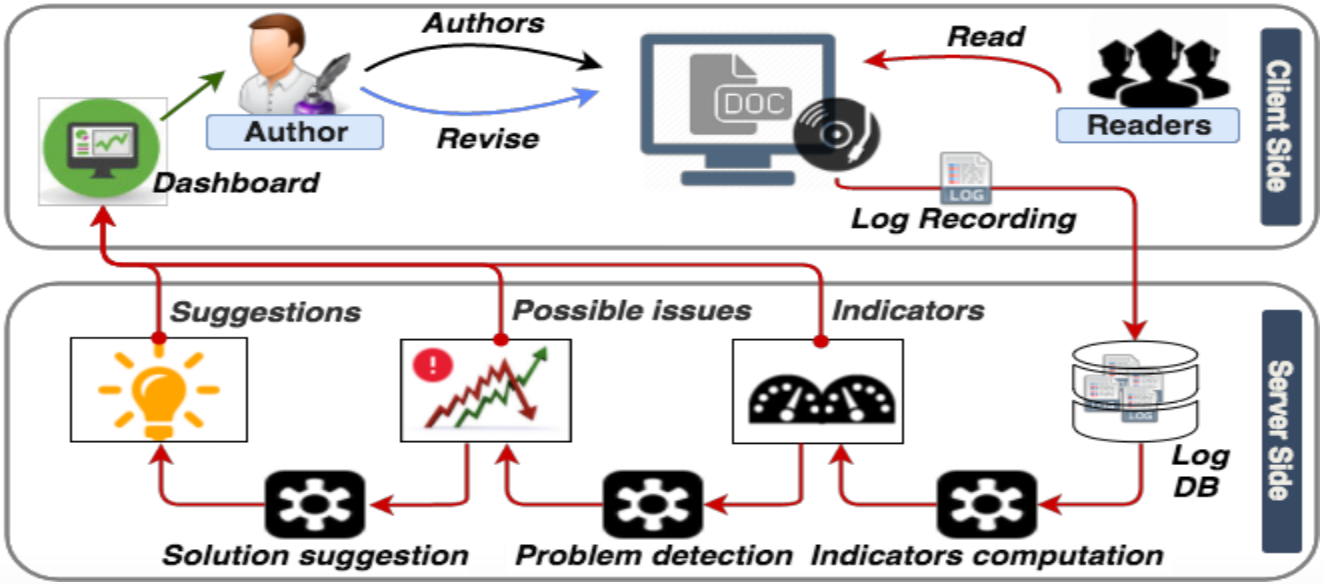
1. Introduction
2. Usage-based document reengineering
3. Course reading analytics
4. CoReaDa
5. Evaluation studies
6. Conclusion

Contributions

1. A general framework for usage-based document reengineering
2. A model of document structures, and the related factors of comprehension and associated issues.
3. Taxonomy of document reengineering actions
4. A reading analytics approach for course revision
5. Reading session concept and algorithm
6. Taxonomy of reading session-based indicators, and their use for issue detection
7. Generation of revision suggestions based on the reading issues
8. CoReaDa: a reading analytics dashboard.
9. A set of evaluation and validation studies

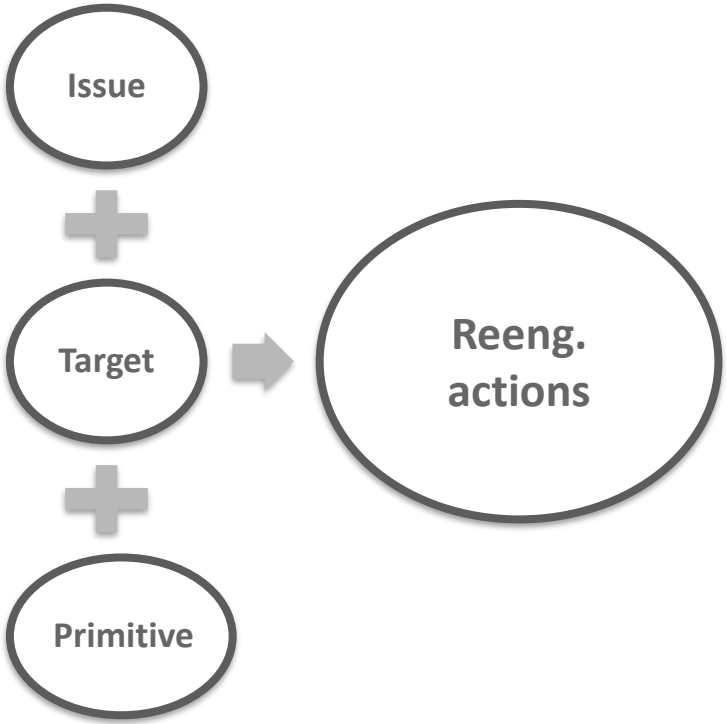
Usage-based document reengineering

"Usage-based document reengineering is the alteration of document content and structures to reconstitute it in a new form, in response to readers' usages"

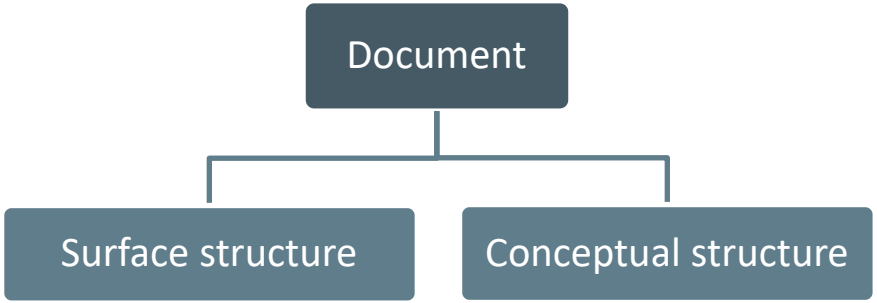


Usage-based document reengineering

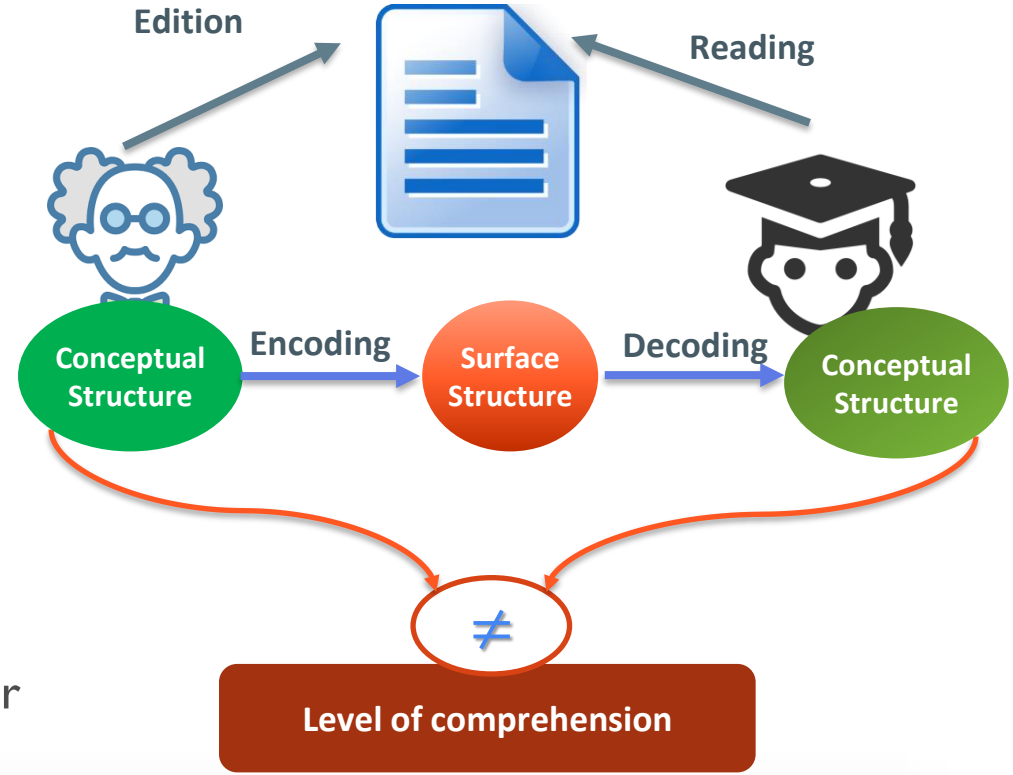
- Reengineering action
  - (edit) primitives
  - (on a) target
  - (to resolve an) issue
- Primitives → Possible edit actions
- Target → Document element
- Issues → from the document structures



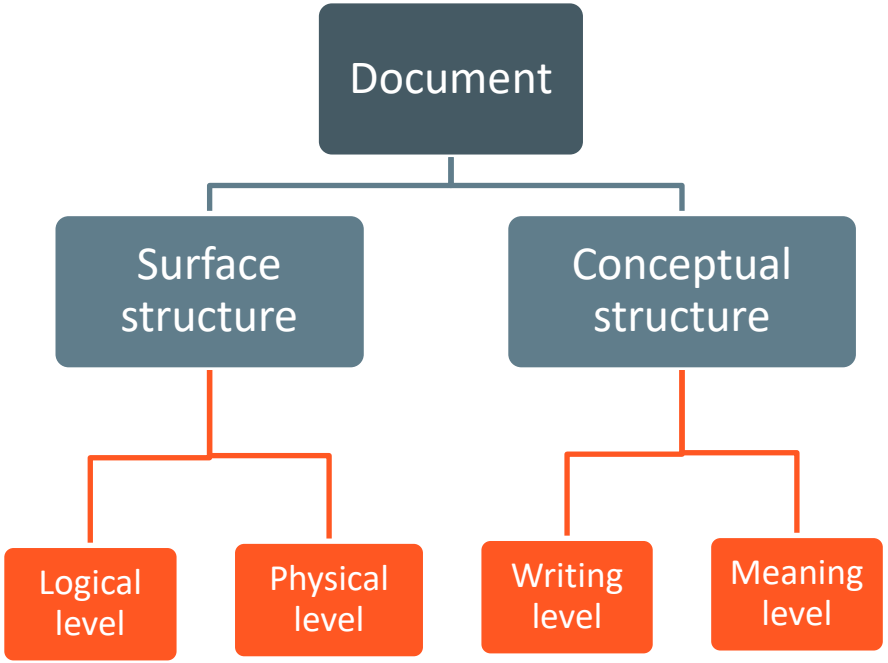
Document model



- The surface structure
  - organization of the document
  - relations between its element
- The conceptual structure
  - what is expressed by the author
  - and how it is expressed



Issues related to document structures

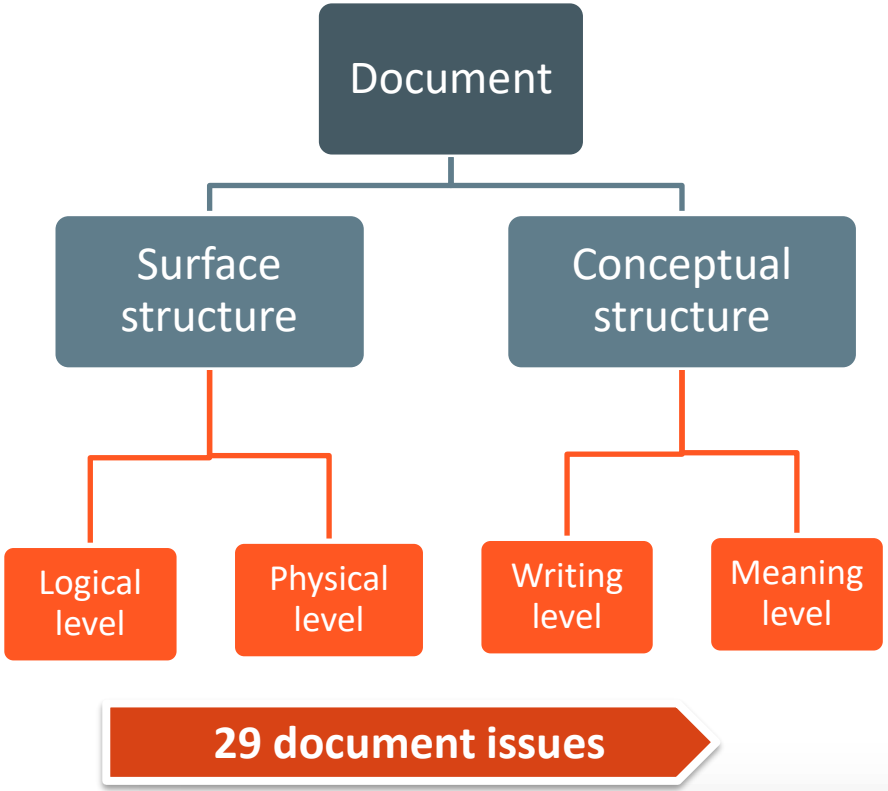


29 document issues

Physical level

1. *Placement on the layout*
- *PL1* Bad location
- *PL2* Inadequate size
2. *Timing and synchronization*
- *PL3* Inadequate temporal inf
- *PL4* Bad synchronization
3. *Linking and navigation*
- *PL5* Inappropriate/useless link
- *PL6* Needed link missing
- *PL7* Broken link

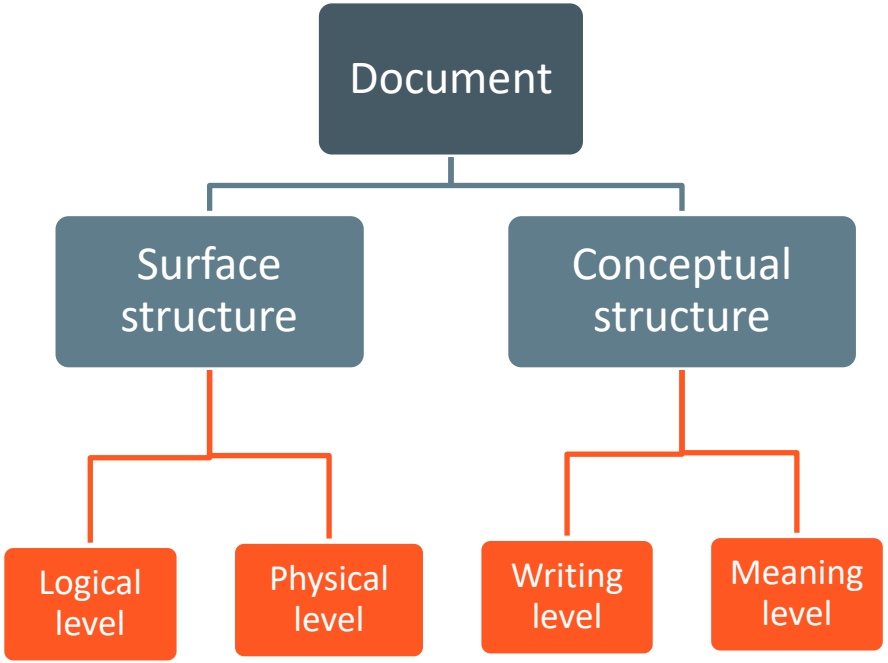
Issues related to document structures



Logical level

1. *Selection of elements*
   - *LL1 Unnecessary/bulky element*
   - *LL2 Non suitable title*
   - *LL3 Element to decompose*
   - *LL4 Element to combine w/others*
2. *Outline and elements sequencing*
   - *LL5 Element not in its best position*
   - *LL6 Late position of the element*
   - *LL7 Early position of the element*

Issues related to document structures



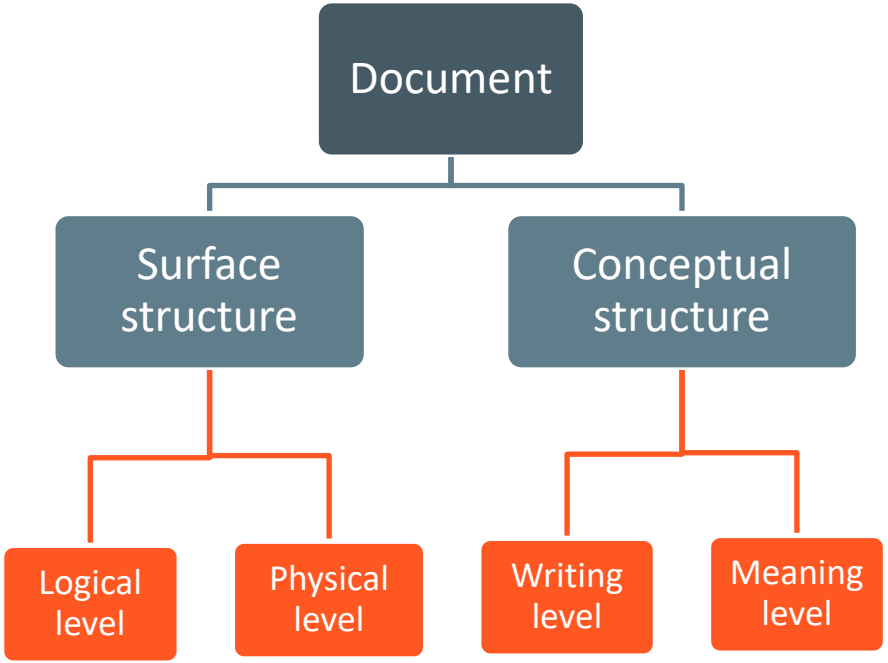
29 document issues

Writing level

1. *Productivity and readability*
   - *WL1* Language and lexical weakness
   - *WL2* Bad syntactic construction

2. Complexity
   - *WL 3* Many new complex information
   - *WL4* Complex construction
   - *WL5* Recall problems

Issues related to document structures



29 document issues

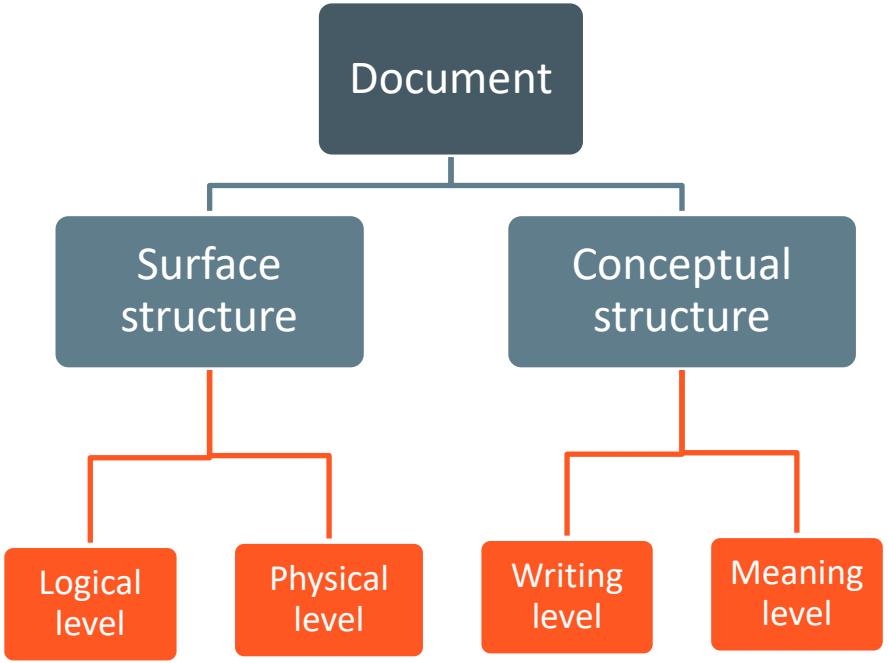
Meaning level

1. *Consistency*
- *ML1* Lack or loss of thematic unit
- *ML2* Contradictions
- *ML 3* Unclear semantic relation

2. *Cohesion*
- *ML4* Unclear connection between ideas
- *ML5* Incoherent ideas

3. *Intentionality and acceptability*
- *ML6* Misunderstanding

Issues related to document structures



29 document issues

Physical level

4. *Informativity*
- *ML7* Marginal or uninformative
- *ML8* Overwhelming

5. *Situationality*
- *ML9* Inadequacy

6. *Intertextuality*
- *ML10* Prerequisites needed

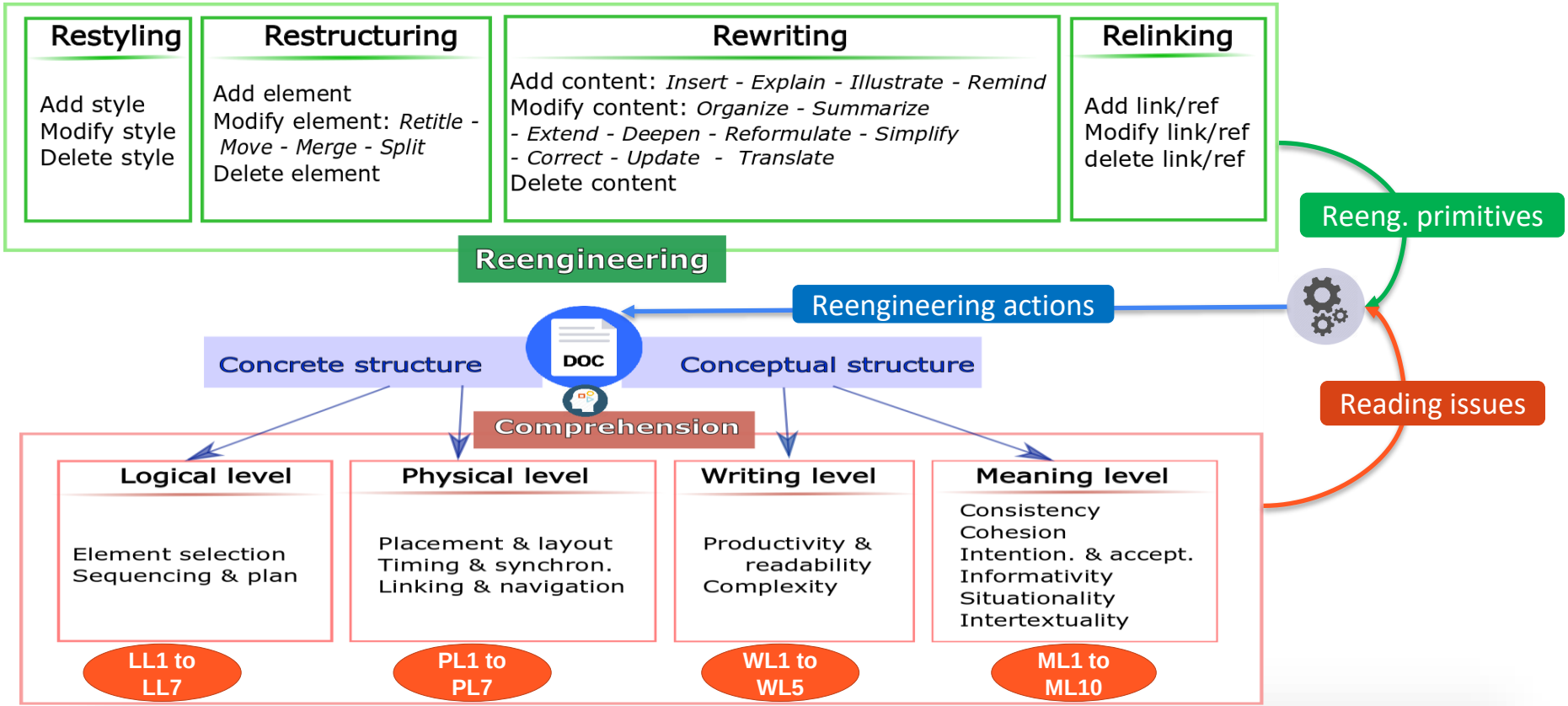
Taxonomy of document reengineering actions

- A primitive
  - impacts a dimension: *style, structure, content,* or *links*
  - produces an effect: *addition, modification* or *suppression*

26 primitives

	Addition	Modification	Deletion
Style	Add rule	Alter rule	Delete rule
Structure	Add element	Retitle – Move – Merge – Split	Delete element
Content	Insert – Explain – Illustrate – Remind	Organize – Summarize – Extend – Deepen – Reformulate – Simplify – Correct – Update –Translate	Delete content
Links	Add ink	Modify link	Del. link

Usage-based document reengineering



Document-related issues and reengineering primitives

Surface structure: Logical Level

Code	Issue	Primitives
LL1	Unnecessary/bulky element	REMOVE
LL2	Non suitable title	RETITLE
LL3	Element to decompose	SPLIT
LL4	Element to combine with others	COMBINE (WITH)
LL5	Element not in its best position	MOVE (TO)
LL6	Late position of the element	MOVE (BACKWARD)
LL7	Early position of the element	MOVE (FORWARD)

Document-related issues and reengineering primitives

Surface structure: Physical Level

Code	Issue	Primitives
PL1	Bad location	MODIFY_STYLE (LOCATION)
PL2	Inadequate size	MODIFY_STYLE(SIZE)
PL3	Inadequate temporal information	MODIFY_STYLE(TIMING)
PL4	Bad synchronization	MODIFY_STYLE(SYNCHRON)
PL5	Inappropriate/useless link	DELETE_LINK
PL6	Needed link missing	ADD_LINK
PL7	Broken link	MODIFY OR DELETE(LINK)

Document-related issues and reengineering primitives

Conceptual structure: Writing Level

Code	Issue	Primitives
WL1	Language and lexical weakness	REFORMULATE AND CORRECT
WL2	Bad syntactic construction	REFORMULATE AND CORRECT
WL3	Many new complex information	REFORMULATE, SUMMARIZE AND CLARIFY
WL4	Complex construction	SPLIT
WL5	Recall problems	REFORMULATE AND CORRECT

Document-related issues and reengineering primitives

Conceptual structure: Meaning Level

Code	Issue	Primitives
ML1	Lack or loss of thematic unit	Update and Correct, Move or Delete
ML2	Contradictions	Update and Correct
ML3	Unclear semantic relationship	Reformulate and Correct, Delete
ML4	Unclear connection between Ideas	Reformulate, Organize , Explain and Extend
ML5	Incoherent ideas	Reformulate, Correct, Explain and Clarify, Move or Delete

Document-related issues and reengineering primitives

Conceptual structure: Meaning Level

Code	Issue	Primitives
ML6	Misunderstanding	Reformulate, Explain, Correct, Clarify, illustrate and Deepen
ML7	Marginal or Uninformative	Deepen, Add, Merge or Delete
ML8	Overwhelming	Split, Clarify, Explain, simplify and Summarize
ML9	Inadequacy	Move or Delete
ML10	Prerequisites needed	Add, Move the element or Add_links

Roadmap

1. Introduction
2. Usage-based document reengineering
3. Course reading analytics
4. CoReaDa
5. Evaluation studies
6. Conclusion

Contributions

1. A general framework for usage-based document reengineering
2. A model of document structures, and the related factors of comprehension and associated issues.
3. Taxonomy of document reengineering actions
4. A reading analytics approach for course revision
5. Reading session concept and algorithm
6. Taxonomy of reading session-based indicators, and their use for issue detection
7. Generation of revision suggestions based on the reading issues
8. CoReaDa: a reading analytics dashboard.
9. A set of evaluation and validation studies

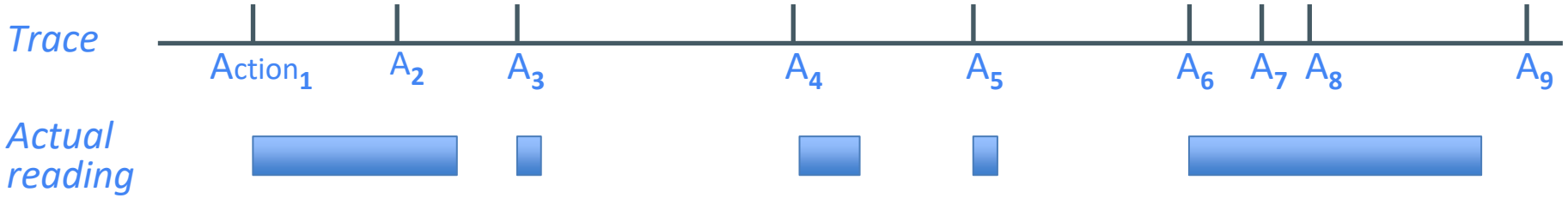
Log-based approach for course revision

- Reengineering courses
  - based on learners' reading activity logs.
  - Sever-based
- A user log = set of his timestamped actions

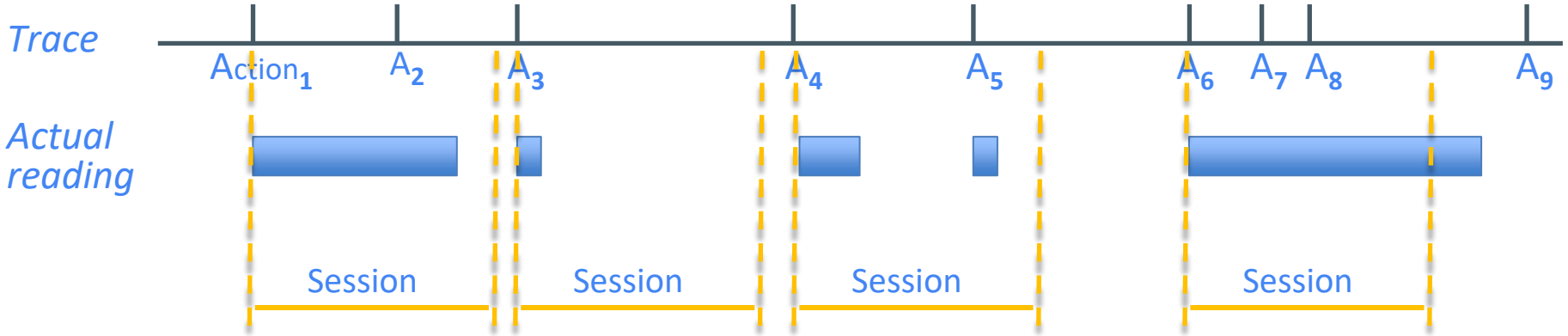
  `<id, user, page, date,…>`
  - Reading activity = set of active + inactive periods
  - Reading session = active period

  → **Need to reconstruct the reading sessions**

Session identification: WUM methods

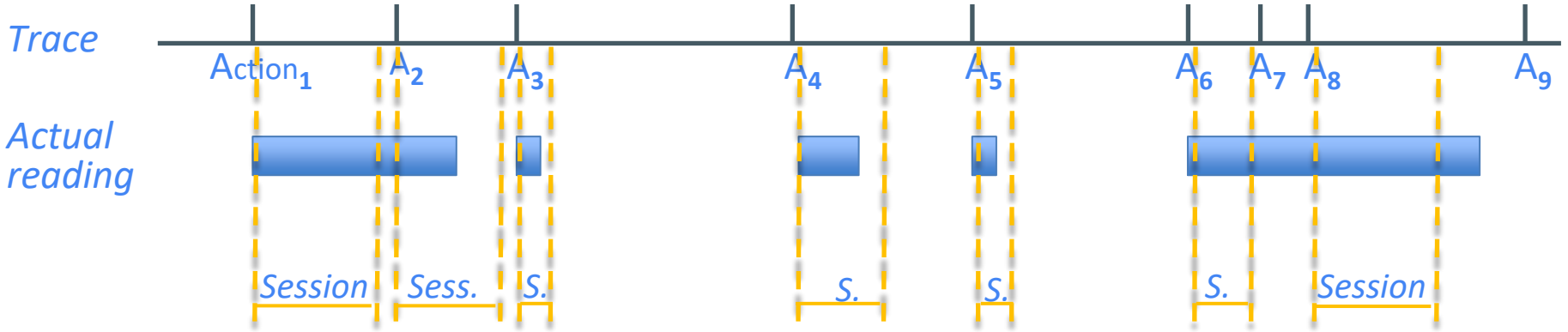


Session identification: WUM methods



- Limit on the total duration of the session
  - Fixed value : 30 min
  - But: cuts continuous activities and merges separate short ones including potential in-between inactive periods

Session identification: WUM methods



- Limit on the page-stay time with a predefined threshold.
  - Fixed value : 10 min
  - But: some pages may be read faster or slower

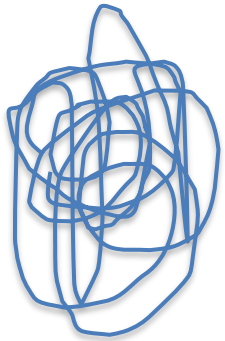
An new algorithm for extracting reading sessions

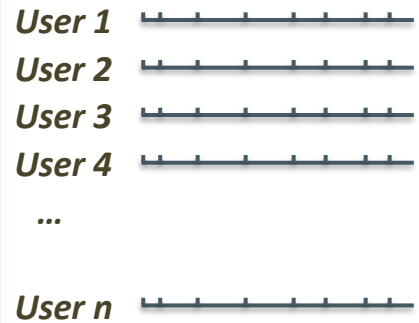
- Dynamic and per-page threshold method
  - Page-stay threshold
  - Each page with its own threshold: page inner-complexity
  - Grounded on data that represent learners' interactions
  - Automatic updating to incoming data and course evolution

A new algorithm for extracting reading sessions

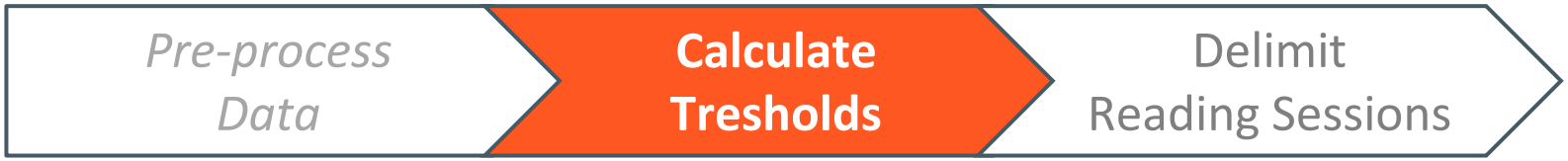
Pre-process Data → Calculate Tresholds → Delimit Reading Sessions

Raw data	*Identify users*	*Compute durations*

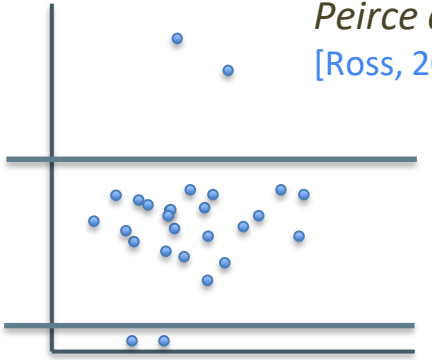




A new algorithm for extracting reading sessions



Eliminate outliers



Compute thresholds values

Elements	Durations	Threshold
Elt 1	$d_{11}, d_{12}, d_{13}, \ldots\ldots$	$=MAX(d_{11},\ldots)$
Elt 2	$d_{21}, d_{22}, d_{23}, \ldots\ldots$	$=MAX(d_{21},\ldots)$
…	……	…
Elt n	$d_{n1}, d_{n2}, d_{n3}, \ldots\ldots$	$=MAX(d_{n1},\ldots)$

A new algorithm for extracting reading sessions

Pre-process Data → Calculate Tresholds → **Delimit Reading Sessions**

Deal with unknown durations

Delimit sessions

Session ends on an action if:

Action duration > time threshold of the element

Reading session-based Indicators

1. Stickiness

Visits ratio

Unique readers ratio

Reading sessions ratio

Reading speed

Interest

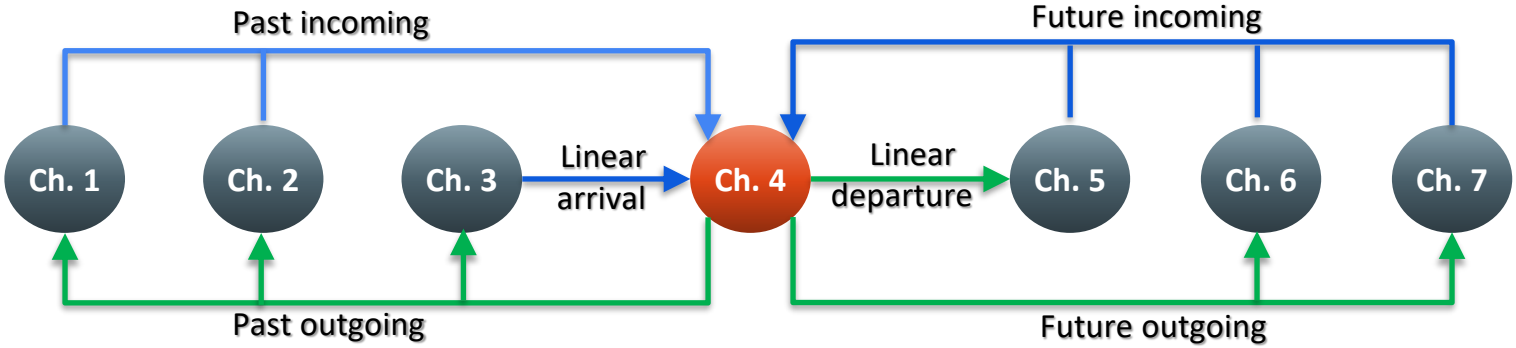
2. Rereading

Rereads ratio

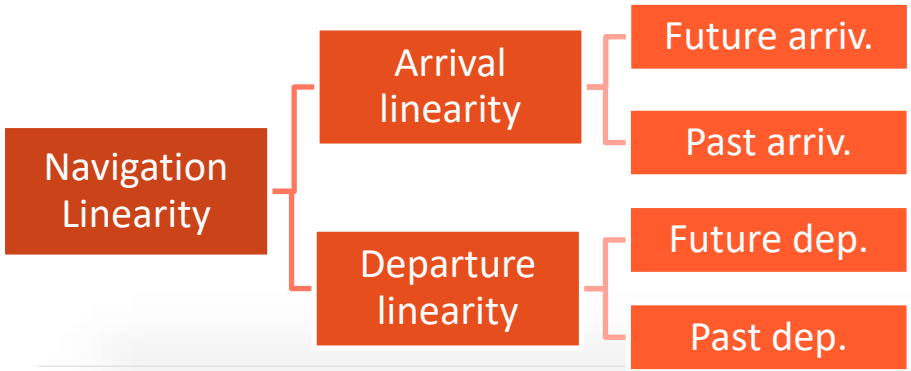
Within-session rereads ratio

Between-session rereads ratio

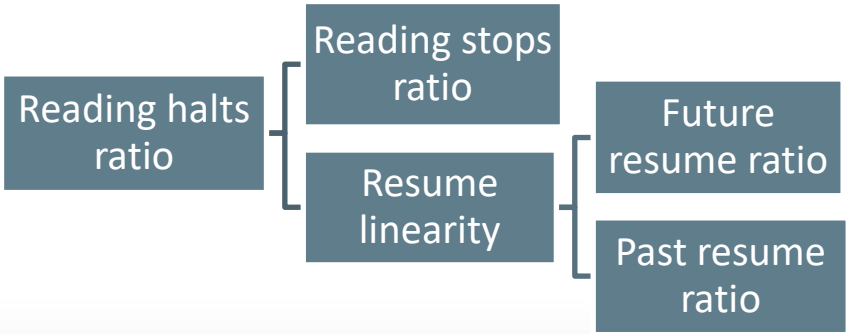
Reading session-based Indicators



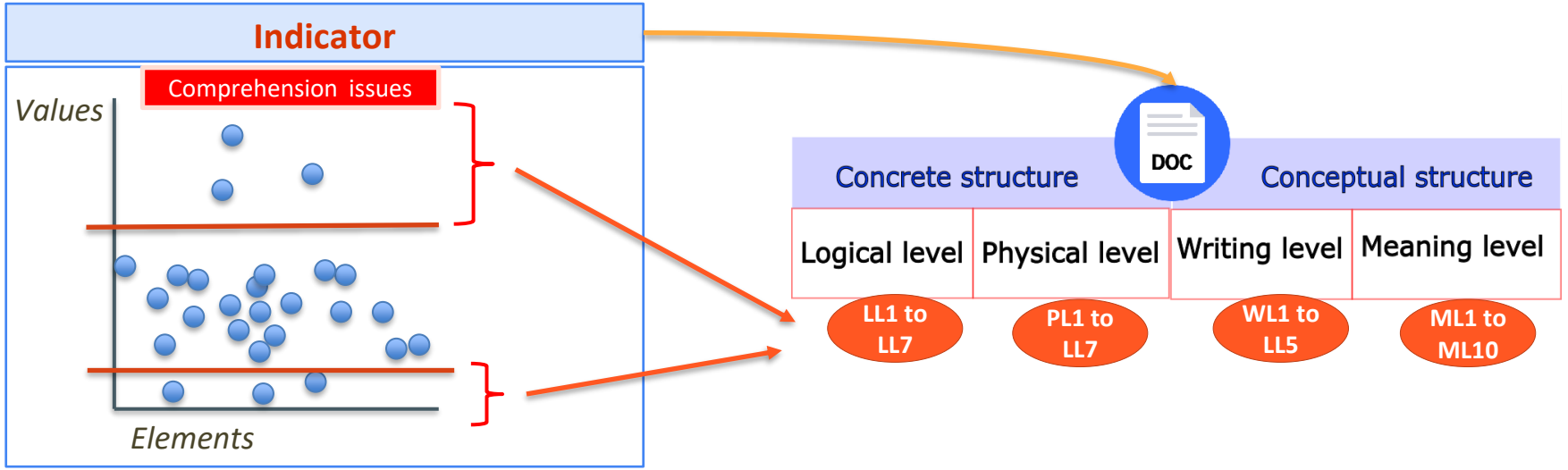
3. Navigation



4. Stop & resume



Indicator-based issue detection & revision suggestion



- Indicators values depend on the construction of the document
- Indicator issues result from document issues
- Revision suggestion = combination of the corresponding primitives

Indicator-based issue detection & revision suggestion

Issue: Very little <u>interest</u>

- Logical level: bad title (LL2)
- Meaning level: uninformative or boring (ML7), out of subject (ML1), not appropriate information (ML9)

Corresponding primitives

- LL2 → (*Restruct.*) Retitle
- ML1 → (*Rewriting*) Update, Correct; (*Restruct.*) Move or Delete
- ML7 → (*Rewriting*) Deepen, Add; (*Restructuring*) Merge or Delete
- ML9 → (*Restruct.*) Move or Delete

Synthetic suggestion: *" If the element is worth presenting on its own: 1) Move it to a more suitable position; 2) and give it a more meaningful and attractive title; and 3) Enrich the element with new content, use graphics and richmedia when possible, and update, correct and deepen the existing content. Otherwise, merge it with an appropriate element or simply delete it"*

Indicator-based issue detection & revision suggestion

Stops	***Issues***	***Revision suggestion***
Multiple reading stops	WL1-WL2 WL3-WL4 ML1 ML3-ML4 ML5-ML10	"If the element is worth presented, move it to a suitable position. Otherwise, merge it with an appropriate one or simply delete it. Also, rewrite the content: enhance its understanding by reformulating and simplifying it, further explaining it and illustrating the ideas. Verify, correct any possible error and update the outdated content."
Multiple nonlinear resumes	LL5-PL5-WL5-ML3-ML4-ML5	"To advocate a linear resume after a halt on the element, move it to a more appropriate position. Facilitate its memorability: reformulate its content, synthesize and clarify the complicated or long parts, and simplify the writing. Also, think to delete some links to/from distant elements from/to this element and replace them with quick reminders where and when needed."
…	…	…

Roadmap

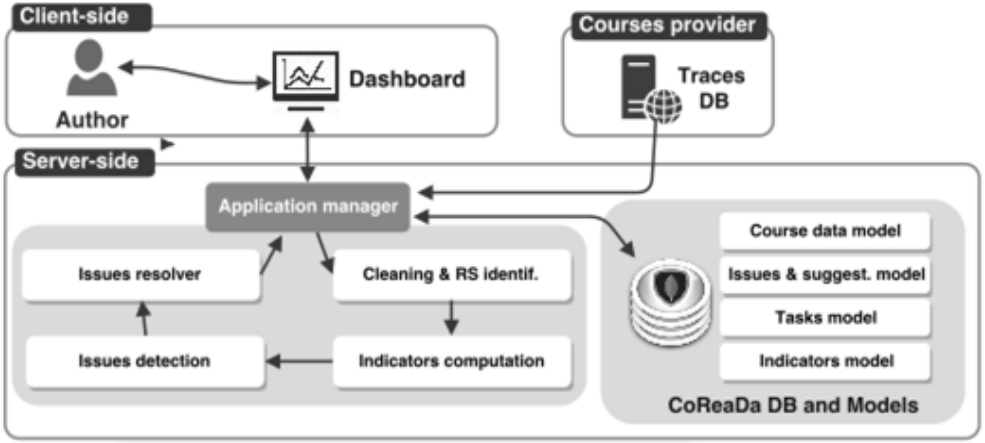
1. Introduction
2. Usage-based document reengineering
3. Course reading analytics
4. CoReaDa
5. Evaluation studies
6. Conclusion

Contributions

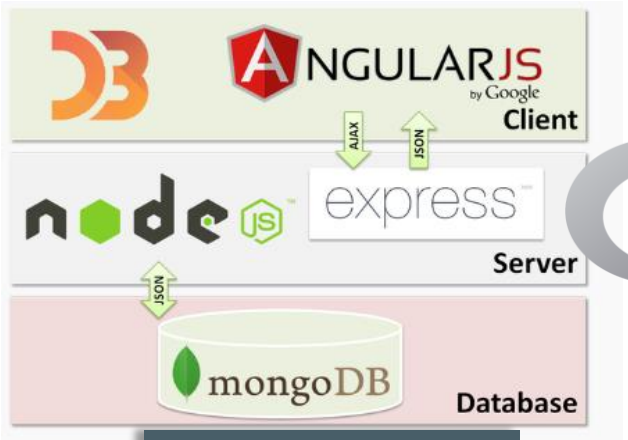
1. A general framework for usage-based document reengineering
2. A model of document structures, and the related factors of comprehension and associated issues.
3. Taxonomy of document reengineering actions
4. A reading analytics approach for course revision
5. Reading session concept and algorithm
6. Taxonomy of reading session-based indicators, and their use for issue detection
7. Generation of revision suggestions based on the reading issues
8. CoReaDa: a reading analytics dashboard.
9. A set of evaluation and validation studies

Implementation of the log-based approach

- Implementation: analytics engine + dashboard
- Dashboard: One-page interface, textual+graphical components



MVC Architecture of CoReaDa

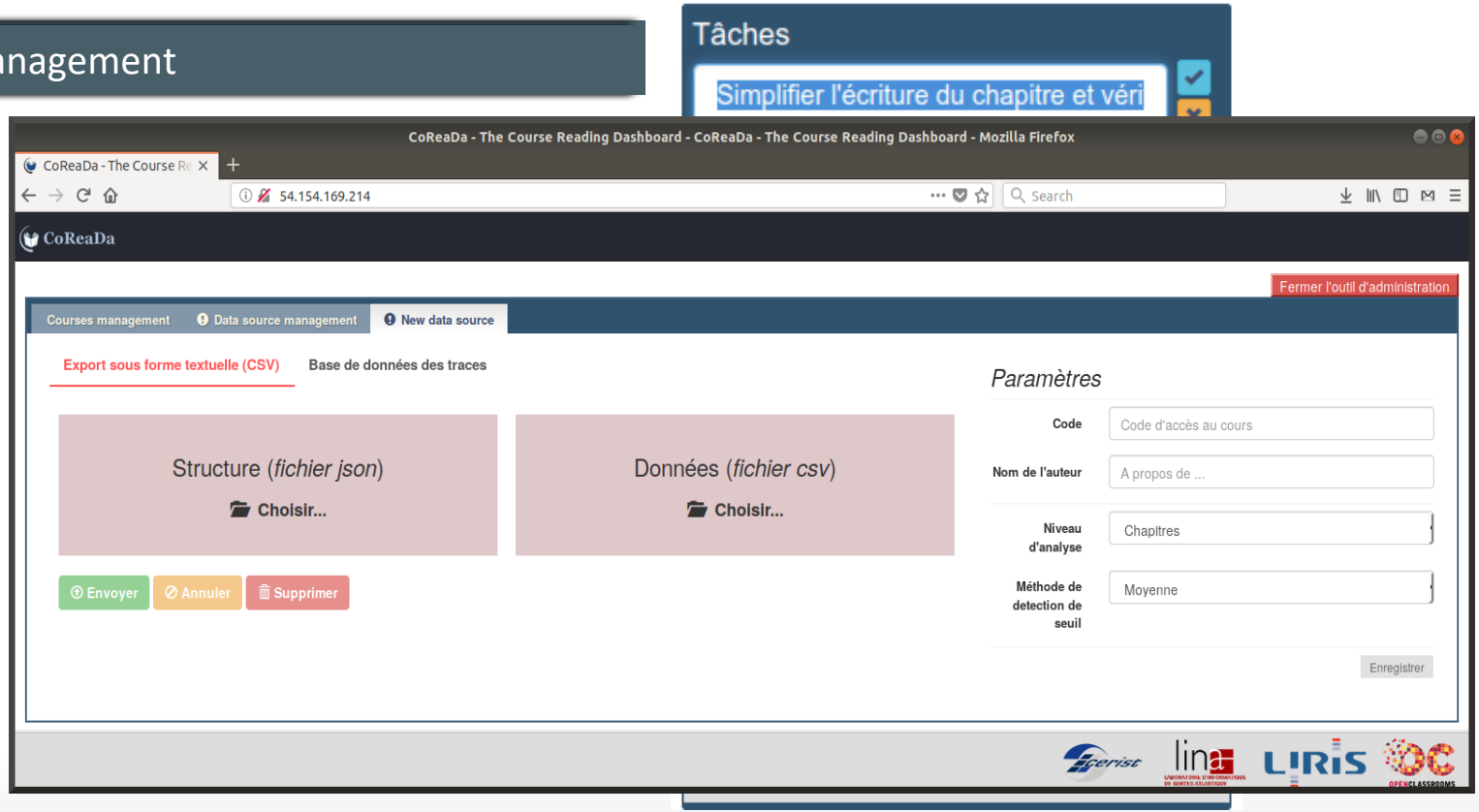


Used technologies

Code: *https://github.com/smadjid/CoReaDa* - Application: *https://bit.ly/CoReaDa*

CoReaDa interface

Courses management



Roadmap

1. Introduction
2. Usage-based document reengineering
3. Course reading analytics
4. CoReaDa
5. Evaluation studies
6. Conclusion

Contributions

1. A general framework for usage-based document reengineering
2. A model of document structures, and the related factors of comprehension and associated issues.
3. Taxonomy of document reengineering actions
4. A reading analytics approach for course revision
5. Reading session concept and algorithm
6. Taxonomy of reading session-based indicators, and their use for issue detection
7. Generation of revision suggestions based on the reading issues
8. CoReaDa: a reading analytics dashboard.
9. A set of evaluation and validation studies

Platform & Objectives

- 5 studies to evaluate our proposals
  - 125 course authors and 26 learners and12 courses Openclassrooms

Study 1: session identification algorithm

1. Compliance with element complexity in terms of size
   - Pearson correlation coefficient element size ⇆ element threshold :

r = 0,82

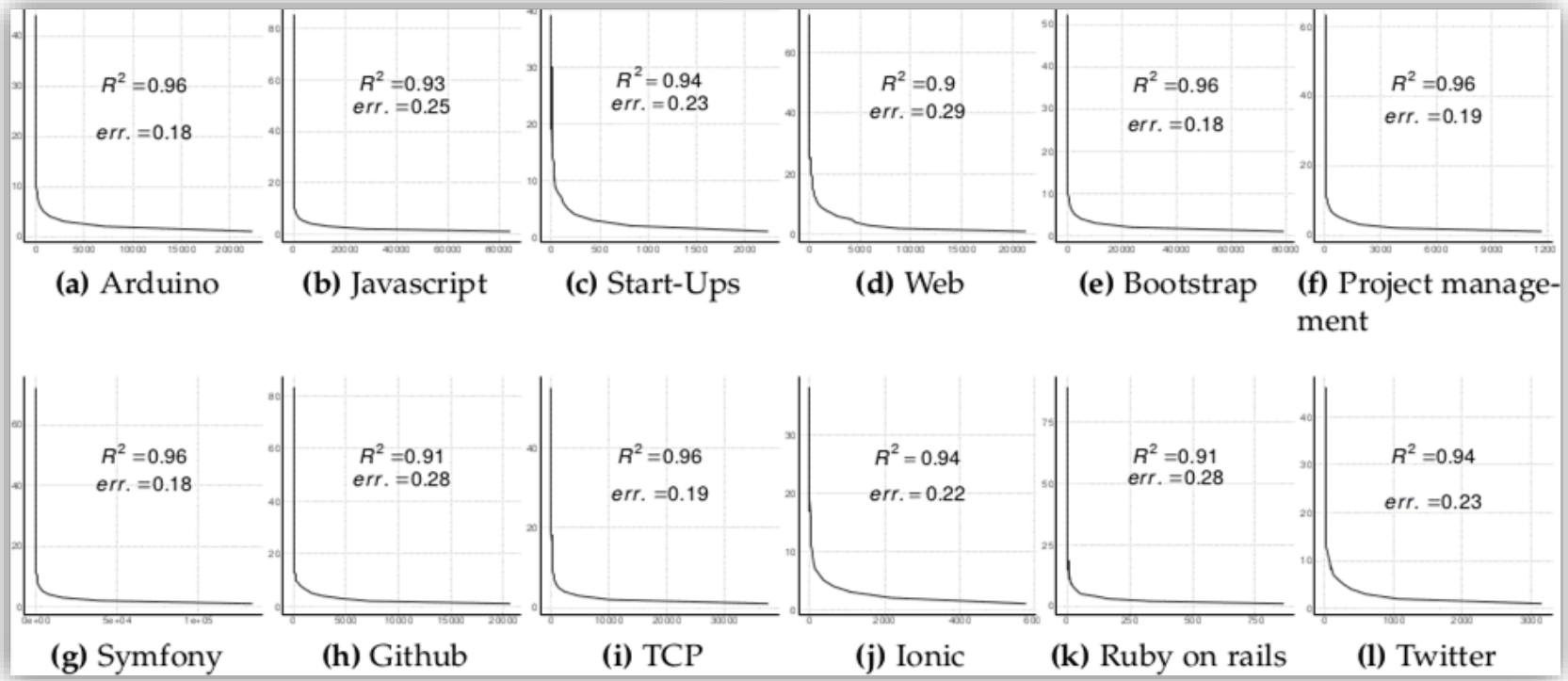
2. Power Law distribution (WUM method):
   - Linear regression: *Log(number of distinct read elements) on Log(total number of reading sessions)*
   - Good results →

regression correlation coefficient R2~1

std. error err~0

Study 1: session identification algorithm

Session size found by the power law distribution



Study 1: session identification algorithm

Comparative results between three session identification methods

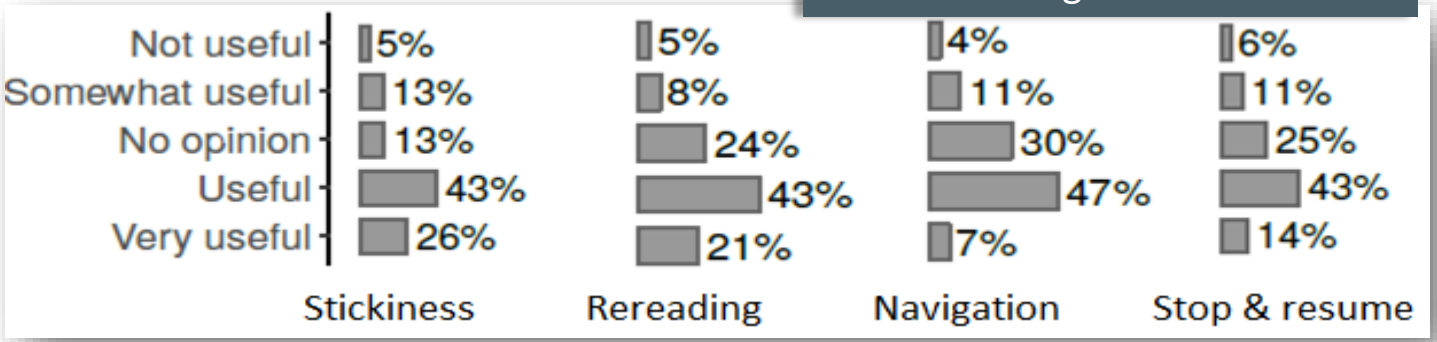
	Reading Session		*10-min Page Thr.*		*30-min Sess. Thres*	
	R^2	*Err*	R^2	*Err*	R^2	*Err*
Bootstrap	**0.96**	**0.18**	0.95	0.22	**0.96**	0.21
Web	**0.90**	**0.29**	0.88	0.31	0.86	0.31
Twitter	**0.94**	**0.23**	**0.94**	0.24	0.91	0.25
Adruino	**0.96**	**0.18**	0.94	0.20	0.93	0.23
JavaScript	**0.93**	**0.25**	0.92	0.26	0.90	0.28
Ionic	**0.94**	0.22	0.93	**0.21**	0.92	0.24
Ruby	0.91	0.28	0.91	0.29	**0.93**	**0.26**
Project	**0.96**	**0.18**	0.92	0.24	0.92	0.23
TCP	**0.96**	**0.19**	0.95	0.20	0.95	0.20
Symfony	**0.96**	**0.29**	0.88	0.31	0.86	0.35
Startups	**0.94**	0.23	0.93	**0.22**	0.93	0.23
Github	0.91	0.28	**0.94**	**0.23**	0.93	0.25

Study 2: Relevance of the set of indicators

- Method
  - online survey: likert scales + free comments
- Participants
  - 125 OpenClassrooms course authors

61% positive - 16% negative – 23% no opinion
Stickiness > Rereading > Stop&resume > Navigation

Authors' ratings of the indicators

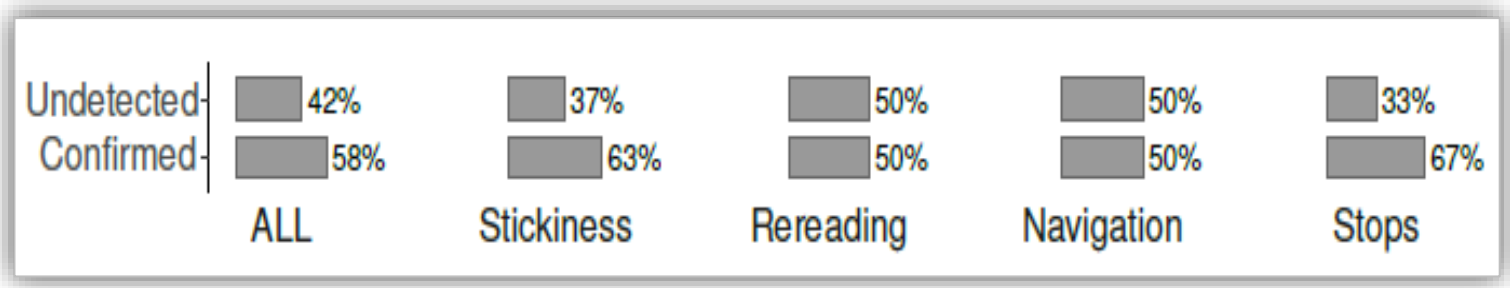


Study 3: Issue detection and revision suggestion

- Participants
  - 12 OpenClassrooms course authors
- Online survey
  - What are your expectations (in terms of problems)?
  - Rate the relevance of the detected issues for course revision (Likert)
  - Rate the relevance of the provided suggestions (Likert)

Study 3: Issue detection and revision suggestion

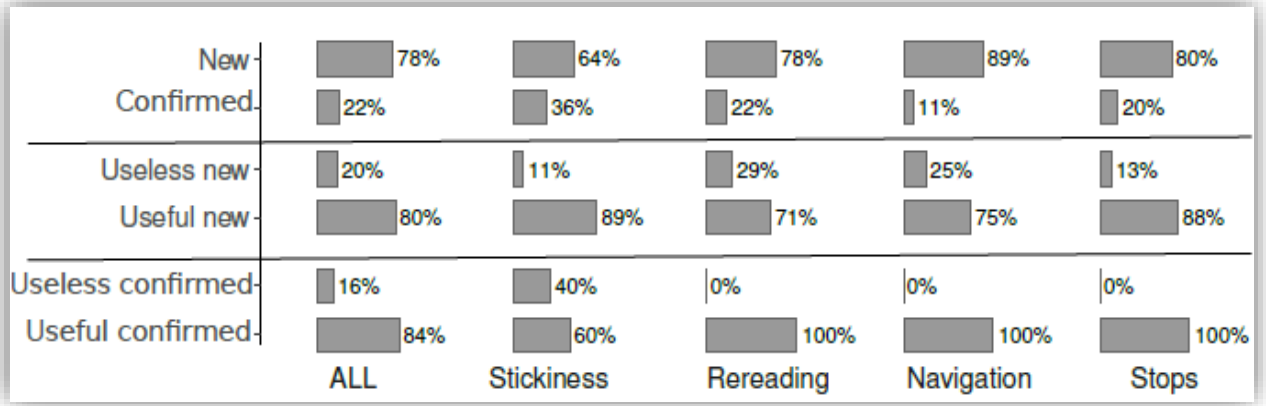
Distribution of the issues expected/detected



42% of authors' beliefs are false
58% of new knowledge

Study 3: Issue detection and revision suggestion

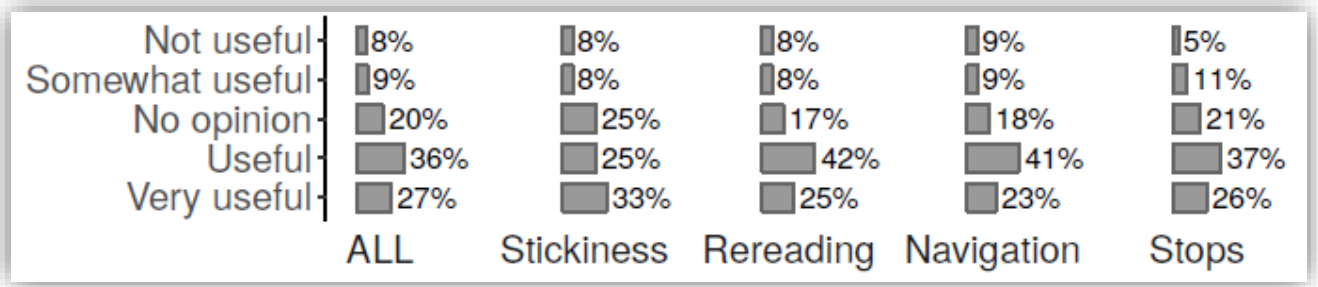
Authors' rating of the relevance of the issues detected



The knowledge is mostly new and is useful

Study 3: Issue detection and revision suggestion

Relevance of the suggestions for course revision



63% positive - 17% negative

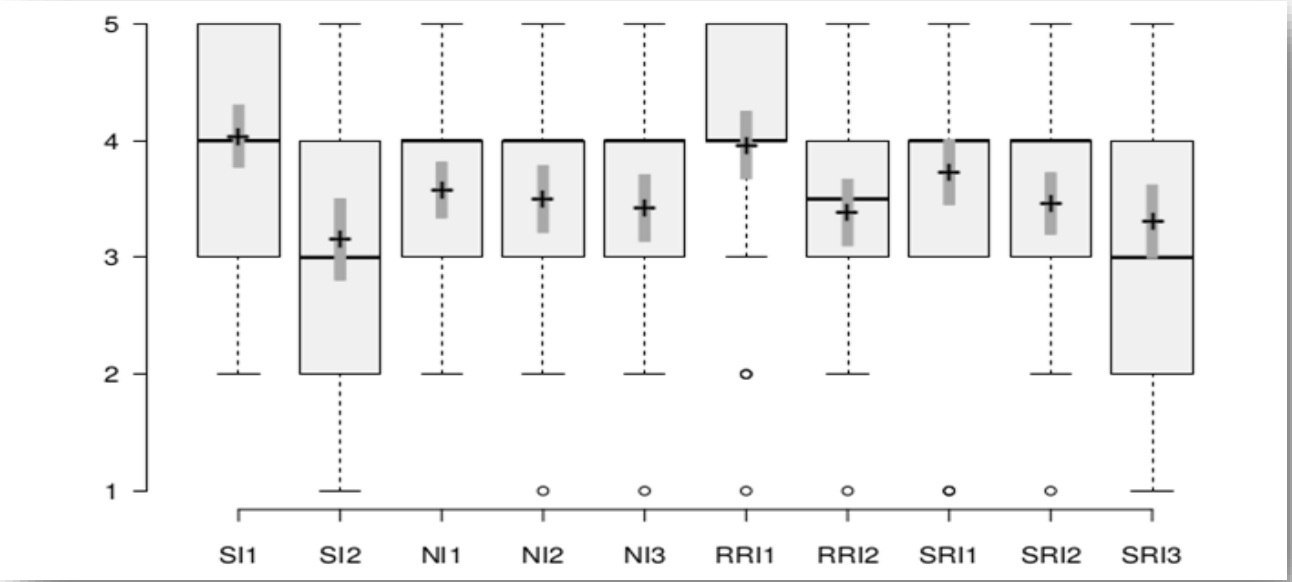
Study 4: Detected issues vs learners' real problems

- 26 learners Openclassrooms
- Subset of 4 courses and 10 types of issues

Cours	*Stickiness*		*Navigation*			*Rereading*		*Stop & resume*			*ALL*
	SI_1	SI_2	NI_1	NI_2	NI_3	RRI_1	RRI_2	SRI_1	SRI_2	SRI_3	
TCP	3	1	2	3	0	1	2	1	0	1	14
Javascript	1	1	1	1	0	1	0	1	0	2	8
Symfony	3	1	2	1	1	2	1	2	1	1	15
Bootstrap	1	1	1	2	1	2	2	0	1	2	13

Study 4: Detected issues vs learners' real problems

Learners' rating of the effectiveness of the issues (1 = very low, 5 = very high)



Almost all ratings >= midpoint 3
All medians & means > midpoint 3

Study 5: a) CoReaDa usability

Task-based experiment

#	*Task purpose*	*Task content*
T1	*Getting help*	Follow the guided tour
T2	*Analyzing the computed indicators*	Find a specific indicator value for a given chapter
T3	*Analyzing the detected issues*	Find a specific issue, review it and mark it as not being an actual problem.
T4	*Performing basic revision tasks*	Select an issue, add the associated suggestion as a task, modify the task and then mark it as done.
T5	*Performing complex operations*	Display all the available indicators and issues to find the chapter(s) with the greatest number of issues

Study 5: a) CoReaDa usability

Performance metrics computed from the tasks results

	Purpose	Success ratio	Avg. Clicks	Avg. Erroneous clicks	Avg. Time (in sec.)
T1	*Getting help*	*100%*	*20*	*0*	*171*
T2	*Analyzing the computed indicators*	*100%*	*6*	*0.7*	*36*
T3	*Analyzing the detected issues*	*100%*	*4.3*	*1.1*	*27*
T4	*Performing basic revision tasks*	*87%*	*7*	*1.6*	*43*
T5	*Performing complex operations*	*75%*	*13*	*3.1*	*89*

***Success ratio**: percent of participants that achieved the task successfully*
***Clicks**: average number of clicks performed to accomplish the task*
***Erroneous clicks**: number of erroneous clicks*
***Time**: average time needed to do the task*

Tasks mostly done quickly and successfully

The guided visit helped!

Study 5: b) CoReaDa acceptance

TAM questionnaire items

	Perceived Ease of Use (PE)
Q1	Learning to use CoReaDa would be easy for me
Q2	I would find it easy to get CoReaDa to revise my course
Q3	My interaction with CoReaDa would be clear and understandable
Q4	I would find CoReaDa to be flexible to interact with
Q5	It would be easy for me to become skillful at using CoReaDa
Q6	I would find CoReaDa easy to use.

	Perceived Usefulness (PU)
Q7	Using CoReaDa would enable me to revise my course more quickly
Q8	Using CoReaDa would improve my revision performance
Q9	Using CoReaDa to revise my courses would increase my productivity
Q10	Using CoReaDa would enhance my effectiveness on course revision
Q11	Using CoReaDa would make it easier for me to revise my courses
Q12	I would find CoReaDa useful in revising my courses

Study 5: b) CoReaDa acceptance

TAM questionnaire results

Perceived Ease of Use			*Perceived Usefulness*		
Item	Mean*	SD	Item	Mean*	SD
Q1	4.38	1.92	*Q7*	4.75	1.83
Q2	5.00	1.60	*Q8*	5.00	1.69
Q3	5.00	1.51	*Q9*	4.75	1.67
Q4	4.88	1.55	*Q10*	5.13	1.81
Q5	5.00	1.51	*Q11*	5.25	1.49
Q6	5.25	1.49	*Q12*	5.50	1.31

**Scale: 1=Strongly disagree to 7=Strongly agree; midpoint : 4*

All means > midpoint 4

PEU + PU → Good attitude to effective use

Roadmap

1. Introduction
2. Usage-based document reengineering
3. Course reading analytics
4. CoReaDa
5. Evaluation studies
6. **Conclusion**

Contributions

1. A general framework for usage-based document reengineering
2. A model of document structures, and the related factors of comprehension and associated issues.
3. Taxonomy of document reengineering actions
4. A reading analytics approach for course revision
5. Reading session concept and algorithm
6. Taxonomy of reading session-based indicators, and their use for issue detection
7. Generation of revision suggestions based on the reading issues
8. CoReaDa: a reading analytics dashboard.
9. A set of evaluation and validation studies

Conclusion

- We proposed a reading analytics approach for detecting learners' needs and assisting authors in revising their courses
- Promising results
  - Good understanding of learners' reading behavior
  - Improved authors' awareness about their course consumption
  - Identification of learners' comprehension issues
  - Generation of revision suggestion
  - Authors support through all the revision process

Perspectives

- A large-scale longitudinal study (in different contexts)
- Traces of other learning activities
- Client-side logs analysis & indicators
- Learners' profiles (background, reading pace) and groups
- Integration of course authoring tools
- Learners' dashboards and tools for self-monitoring

Academic output

- “A framework for usage-based document reengineering” (2013) in Proceedings of the *2013* ACM Symposium on Document Engineering (DocEng’13), Florence, Italy, pages 99–102. ACM
- “Towards reading session-based indicators in educational reading analytics” (2015) in Design for Teaching and Learning in a Networked World. Lecture Notes in Computer Science, vol 9307, pages 297–310. Springer, Cham
- “Leveraging Learners’ Activity Logs for Course Reading Analytics Using Session-Based Indicators” (in press) in International Journal of Technology Enhanced Learning (IJ-TEL). Inderscience
- “Towards fine-grained reading dashboards for online course revision”. Educational Technology Research and Development (ETR&D), (in press). Springer

Thank you for your attention!