Facilitating Discourse Analysis with Interactive Visualization
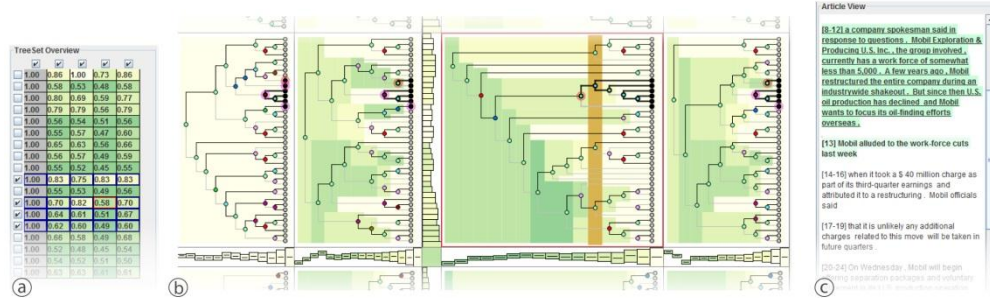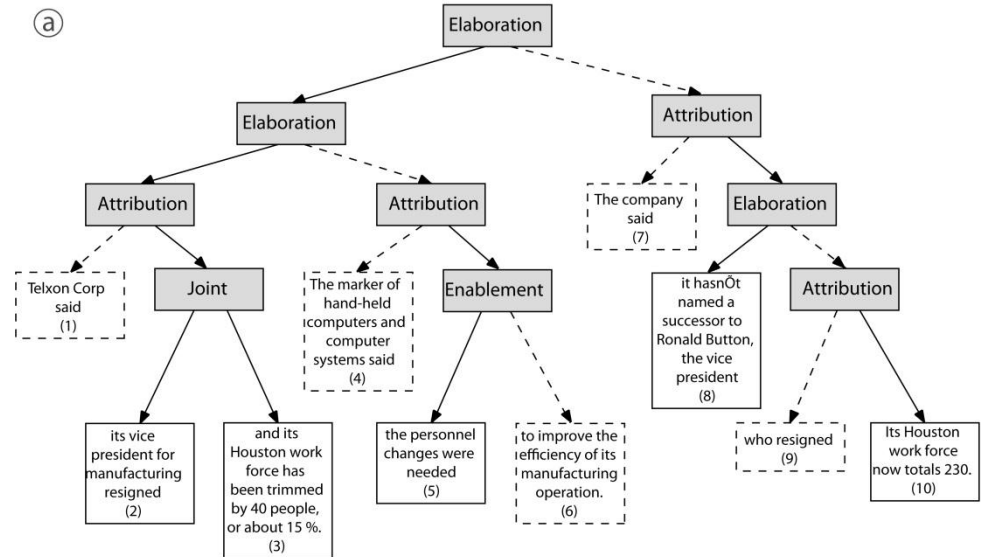
Jian Zhao

Fanny Chevalier

Christopher Collins

Ravin Balakrishnan

UNIVERSITY OF TORONTO

UOIT
CHALLENGE INNOVATE CONNECT
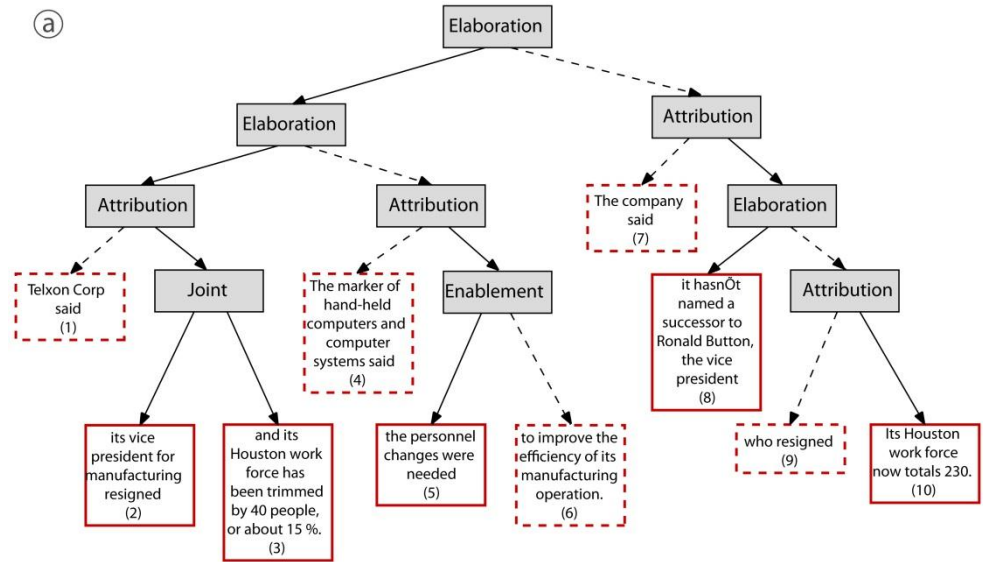
dgp
dynamic graphics project

Discourse Analysis of Text

Telxon Corp. said its vice president for manufacturing resigned and its Houston work force has been trimmed by 40 people, or about 15%. The maker of hand-held computers and computer systems said the personnel changes were needed to improve the efficiency of its manufacturing operation. The company said it hasn't named a successor to Ronald Bufton, the vice president who resigned. Its Houston work force now totals 230.
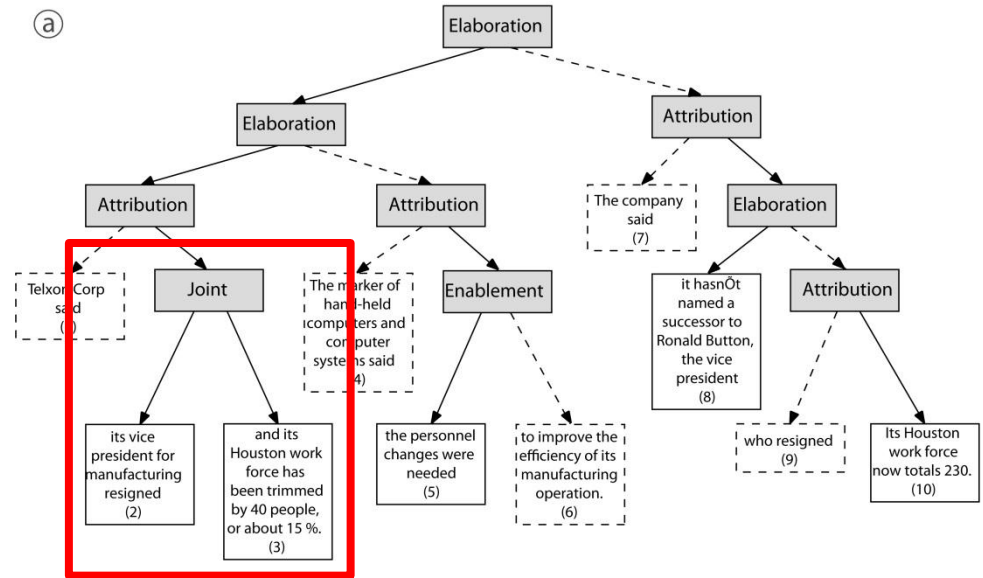

2

DISCOURSE PARSER

Telxon Corp. said / its vice president for manufacturing resigned / and its Houston work force has been trimmed by 40 people, or about 15%. / The maker of hand-held computers and computer systems said / the personnel changes were needed / to improve the efficiency of its manufacturing operation. / The company said it hasn't named a successor to Ronald Bufton, the vice president / who resigned. / Its Houston work force now totals 230.
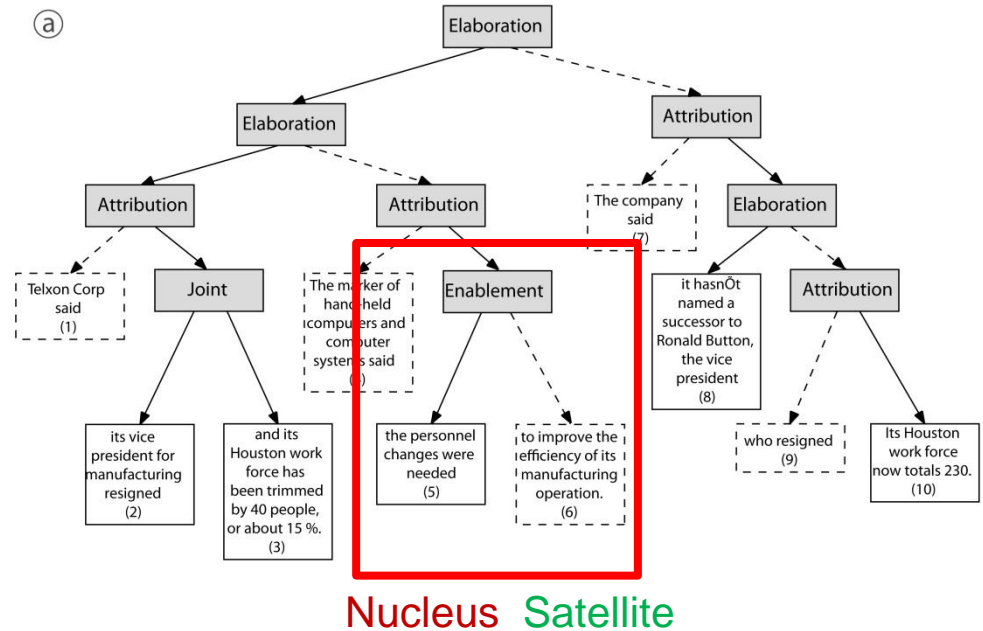
Leaf nodes are EDUs

3

Discourse Parser

Telxon Corp. said **/** its vice president for manufacturing resigned **/** and its Houston work force has been trimmed by 40 people, or about 15%. **/** The maker of hand-held computers and computer systems said **/** the personnel changes were needed **/** to improve the efficiency of its manufacturing operation. **/** The company said it hasn't named a successor to Ronald Bufton, the vice president **/** who resigned. **/** Its Houston work force now totals 230.

Nucleus   Nucleus

Merging nodes and labeling *Joint* relation

Discourse Parser

Telxon Corp. said **/** its vice president for manufacturing resigned **/** and its Houston work force has been trimmed by 40 people, or about 15%. **/** The maker of hand-held computers and computer systems said **/** the personnel changes were needed **/** to improve the efficiency of its manufacturing operation. **/** The company said it hasn't named a successor to Ronald Bufton, the vice president **/** who resigned. **/** Its Houston work force now totals 230.

Nucleus Satellite

Merging nodes and labeling *Enablement* relation

Discourse Parser

Telxon Corp. said / its vice president for manufacturing resigned / and its Houston work force has been trimmed by 40 people, or about 15%. / The maker of hand-held computers and computer systems said / the personnel changes were needed / to improve the efficiency of its manufacturing operation. / The company said it hasn't named a successor to Ronald Bufton, the vice president / who resigned. / Its Houston work force now totals 230.

ⓐ
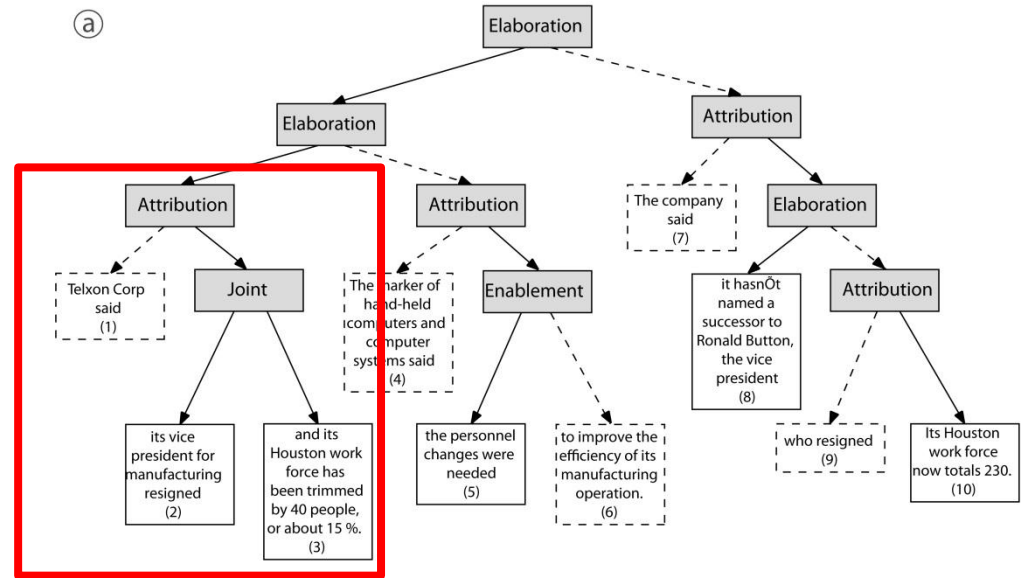
Satellite Nucleus

Merging nodes and labeling *Attribution* relation

6

RESOURCES

- RST Discourse Treebank Corpus

  - 385 Wall Street Journals
  - Annotated manually by linguists
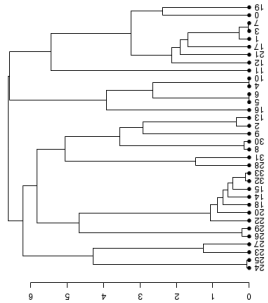- Different parsing algorithms

To Design Better Parsers

- Meaningful text chunks merging on each tree level?

- Errors propagations during the node merging and relation labelling process?

- Differences of parsing results across algorithms?
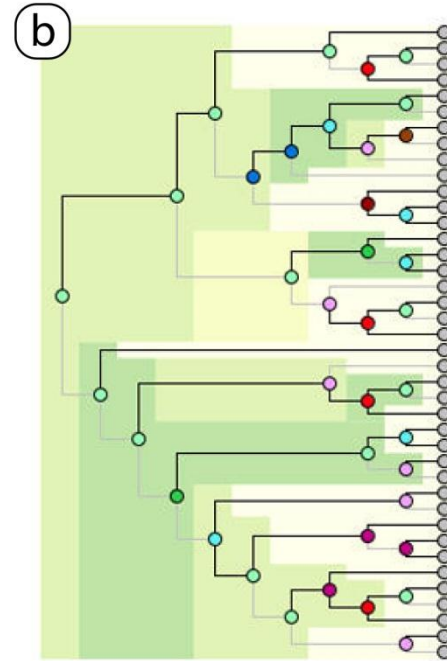
- Differences of parsing results across documents?

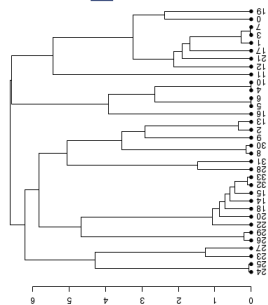# What in their minds vs. What they really see
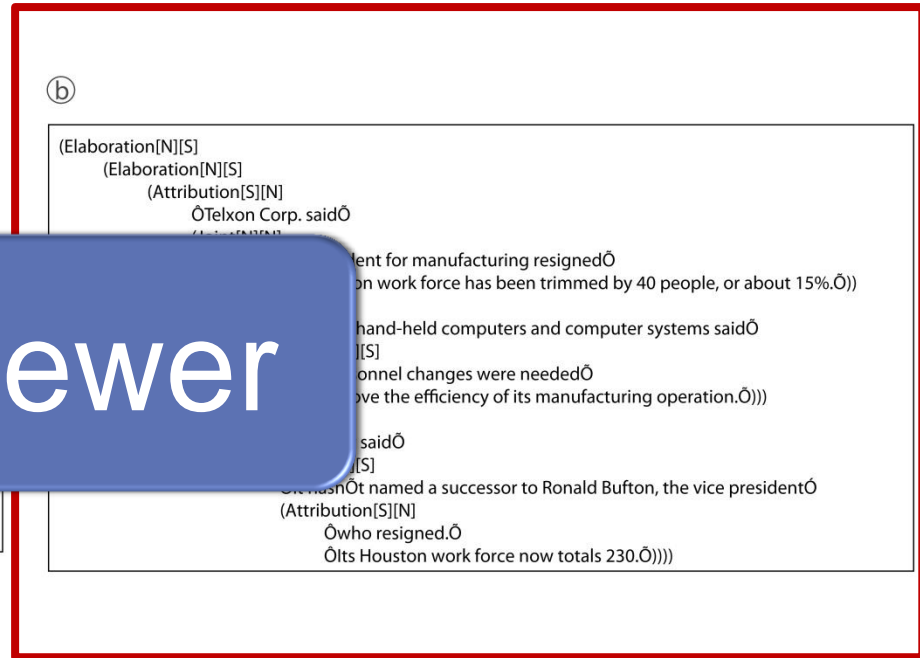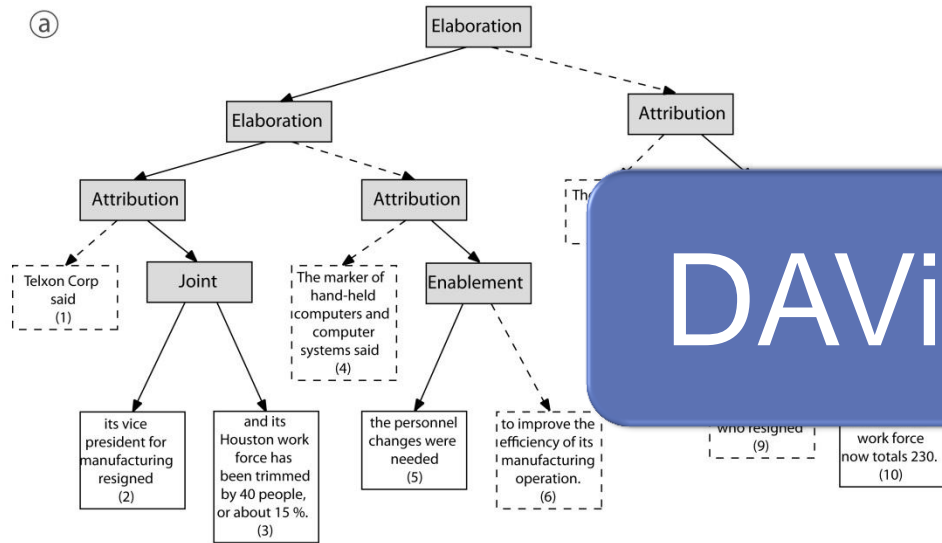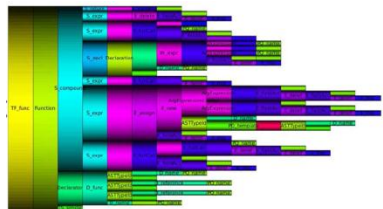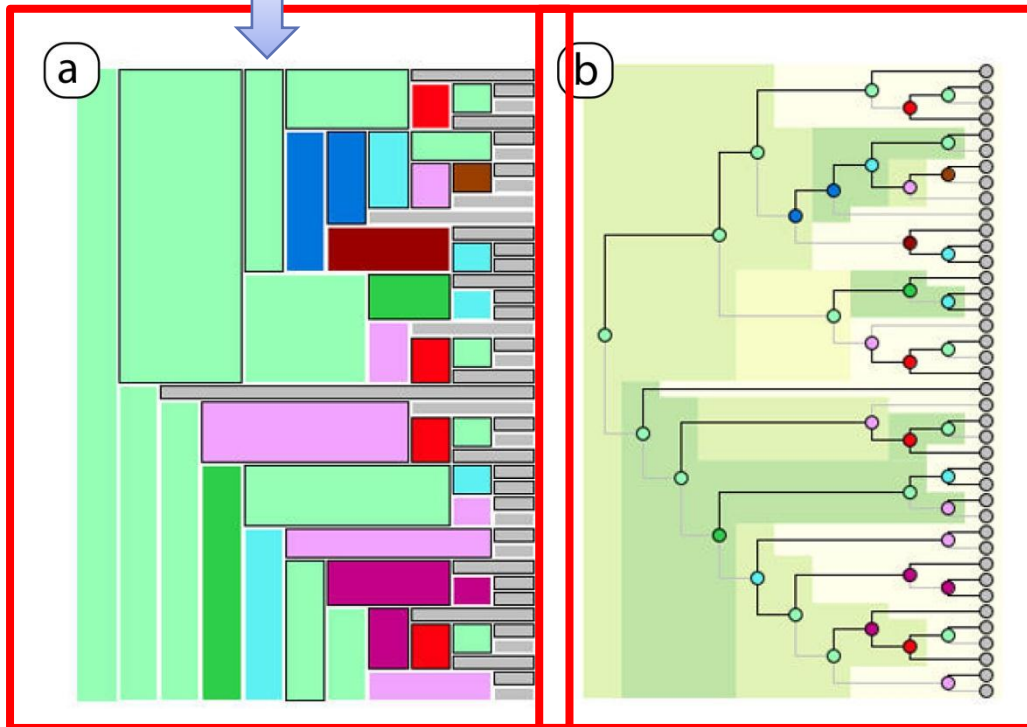
DAViewer: Core Visualizations

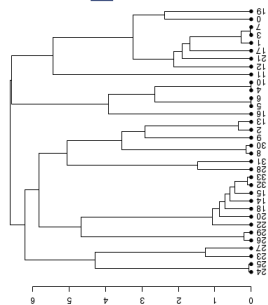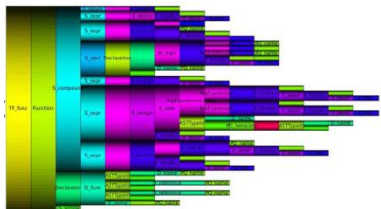dendrogram

DAVIEWER: CORE VISUALIZATIONS

Icicle-plot

+

dendrogram

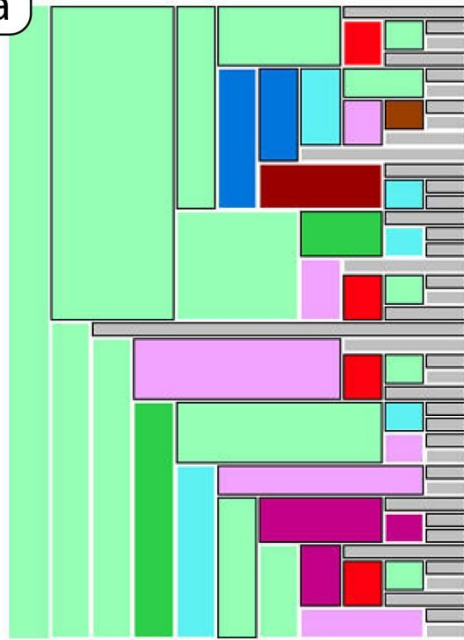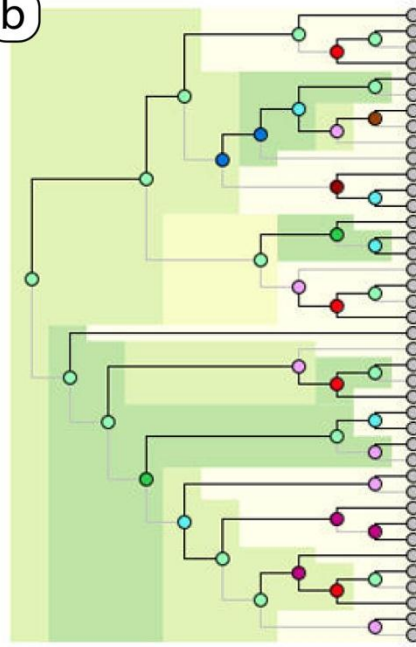# DAViewer: Core Visualizations

Icicle-plot

+

dendrogram
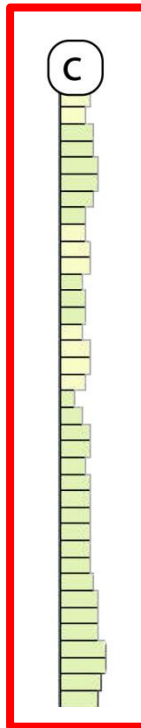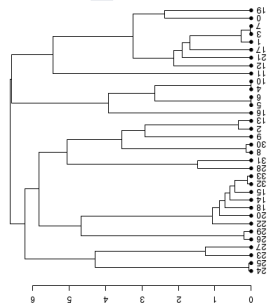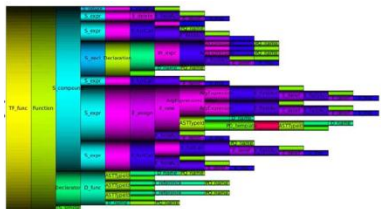
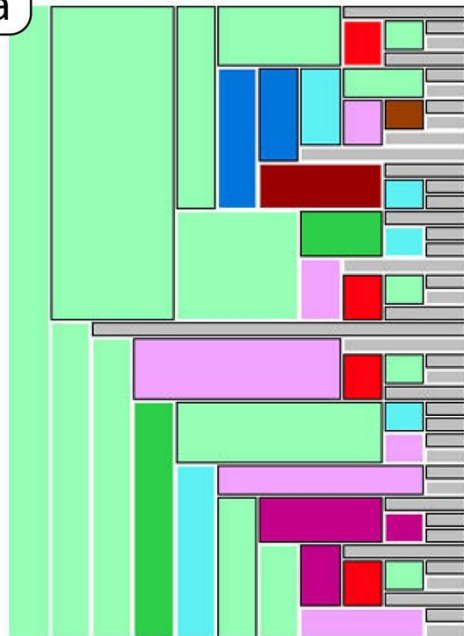# DAViewer: Core Visualizations

Icicle-plot

+

dendrogram
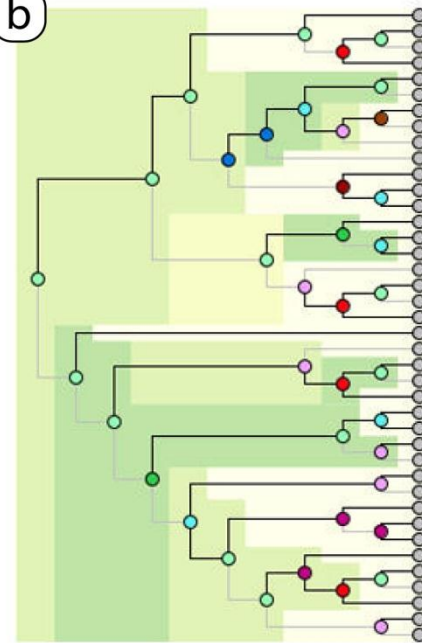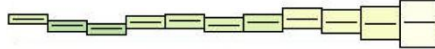
DAVIEWER: INTERFACE

DESIGN DAVIEWER

Problem identification and quick prototyping

Initial implementation

Refined prototype

System release and field test
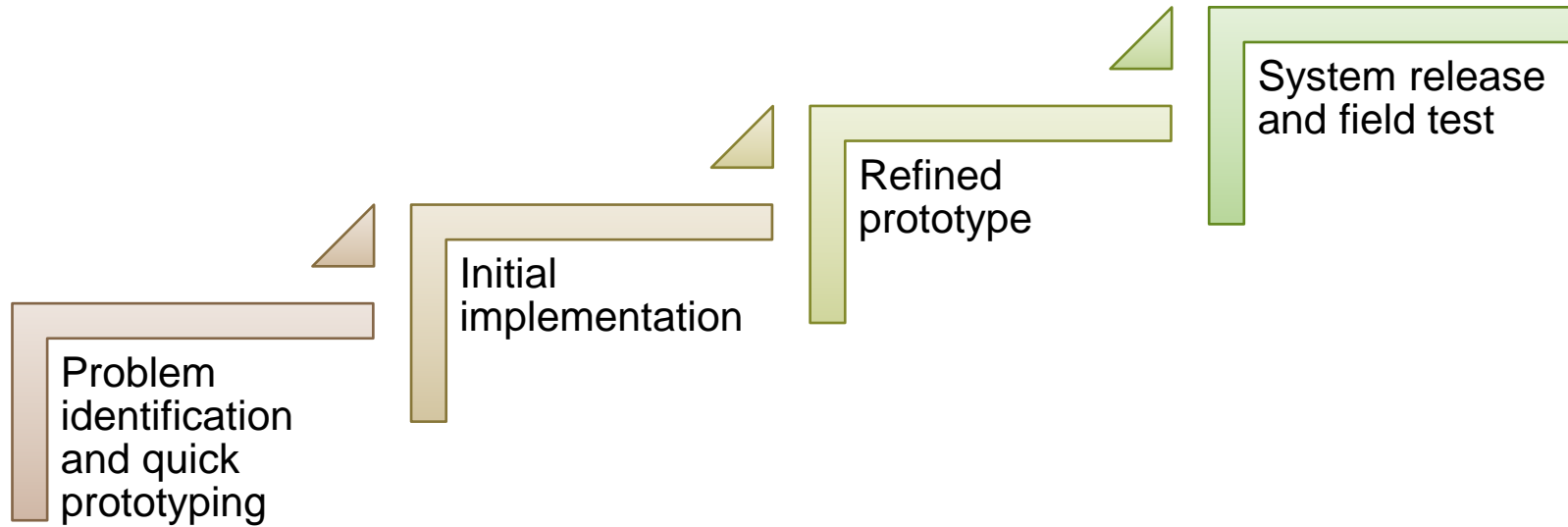
Evaluation

- 3 weeks in realistic work settings

- 1 domain expert user

- Interaction logging

- Analysis diaries

- Email and phone communication

- Interviews

RESULTS

- Log files analysis

- Qualitative feedbacks

- Usage scenarios

RESULTS

- **Log files analysis**

- Qualitative feedbacks

- Usage scenarios

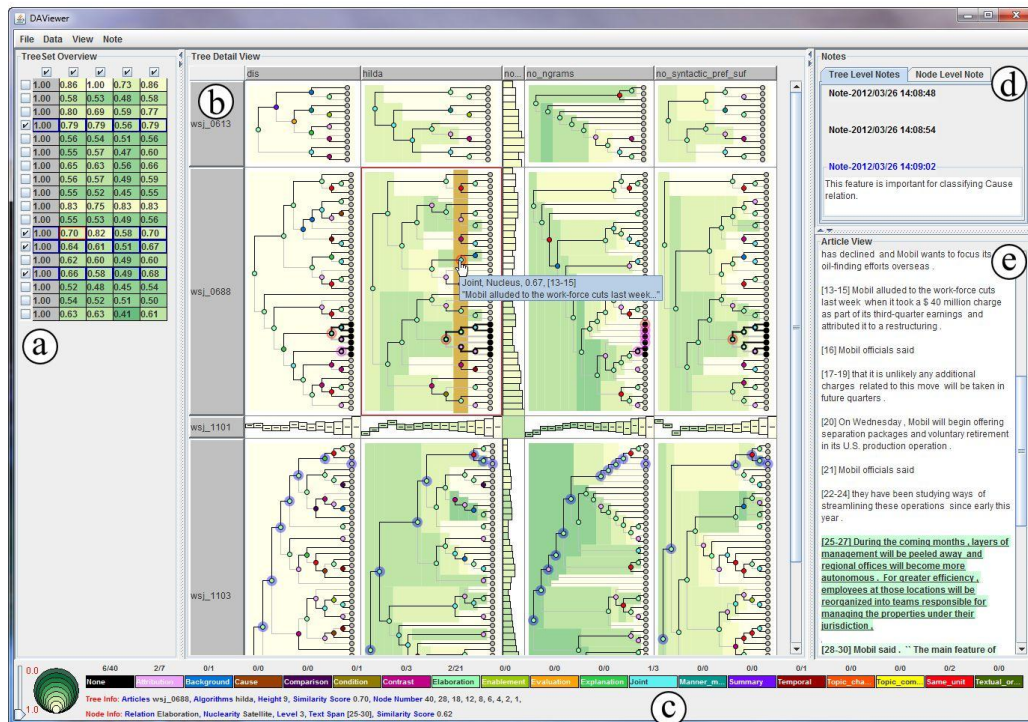Overall function usage

ⓒ

dataset management (0.5%)

overview and tree subset (9.2%)

detail panel manipulation (14.8%)

tree exploration (53.2%)

text display (0.3%)

annotation (6.4%)

querying (9.8%)

interactive legend filtering (5.8%)
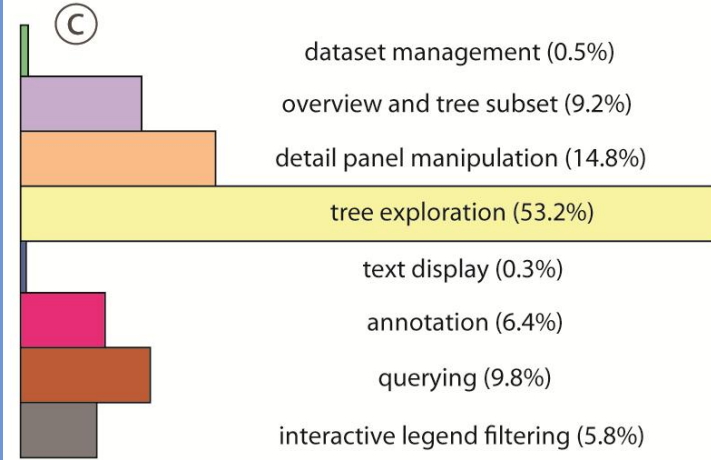
Function usage in a typical session

ⓐ

Time

Results

- Log files analysis

- **Qualitative feedbacks**

- Usage scenarios

"I used to insert debug code inside the parsers, for a verbose output, but it is tedious and almost impossible to find patterns. Now, I can do it visually by just creating what I need and clicking a button."

"I know I want the tree levels aligned from the bottom, but I can't draw such layout from the output file in one round since the tree nodes are indented in the depth-first order."
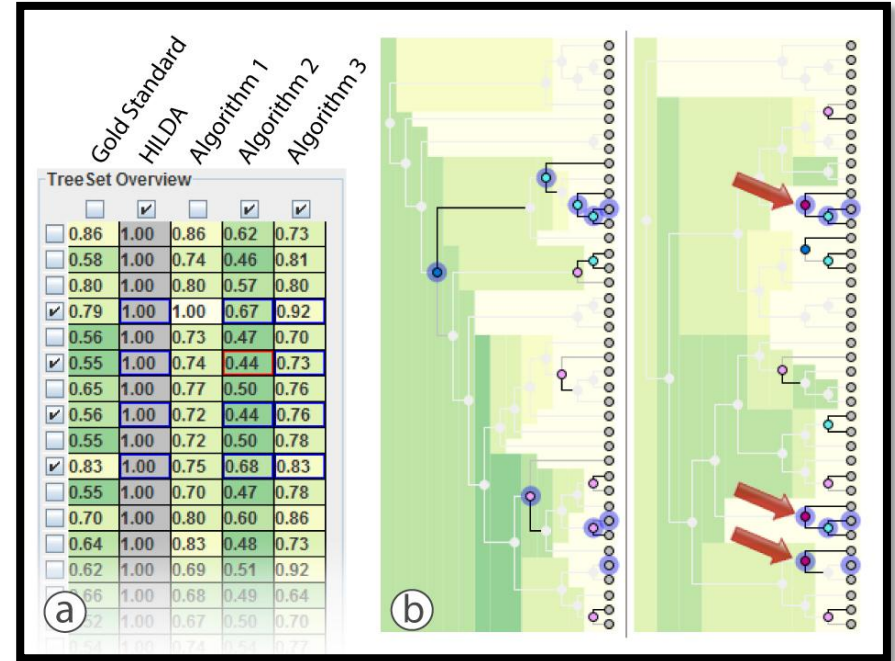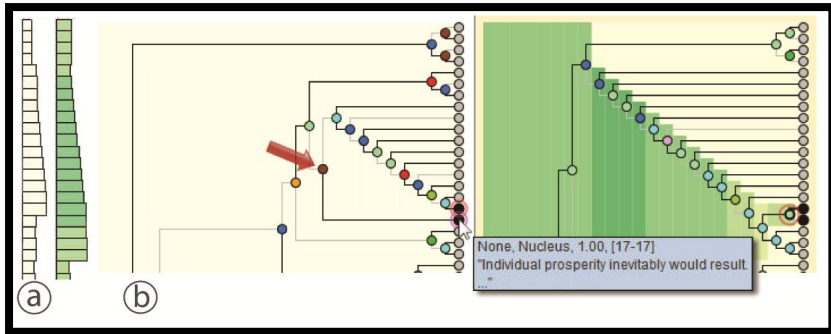
"I used to draw trees by hand according to the output text files, so I could only compare 2 or 3 trees at a time and they are basically simple trees around 5 levels. With DAViewer I can compare many large trees efficiently. All trees are nicely aligned and the interactions allow me to focus on subtrees easily."
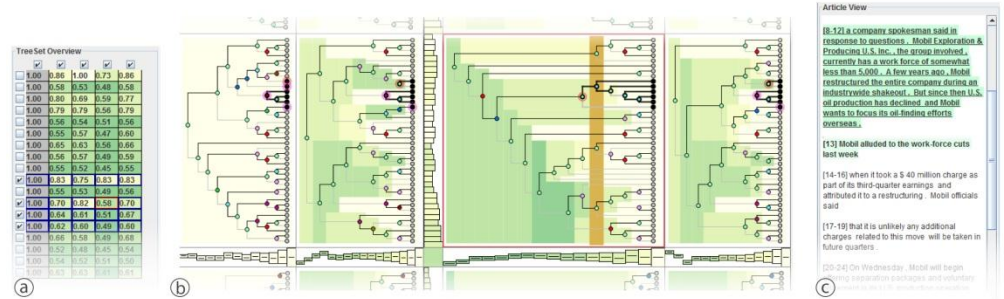
RESULTS

- Log files analysis

- Qualitative feedbacks

- **Usage scenarios**

FUTURE WORK

- More evaluation for widespread adoption

- More interactions for tree exploration

- More integrated working environment

- Extension of visualization techniques to other domains for tree structure comparison and analysis

DAViewer

FACILITATING DISCOURSE ANALYSIS WITH INTERACTIVE VISUALIZATION

Jian Zhao
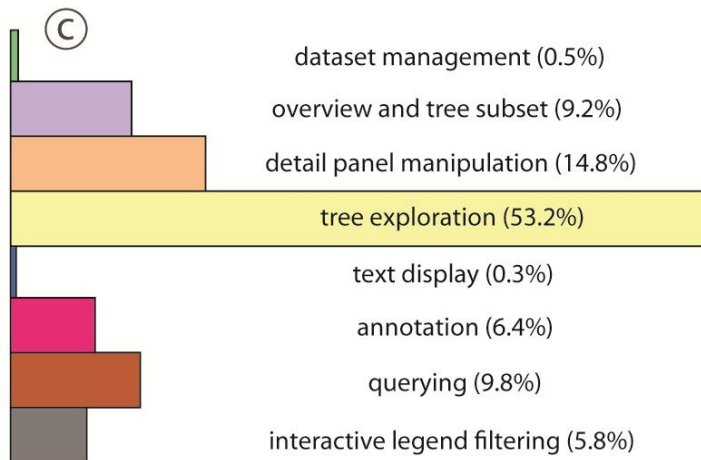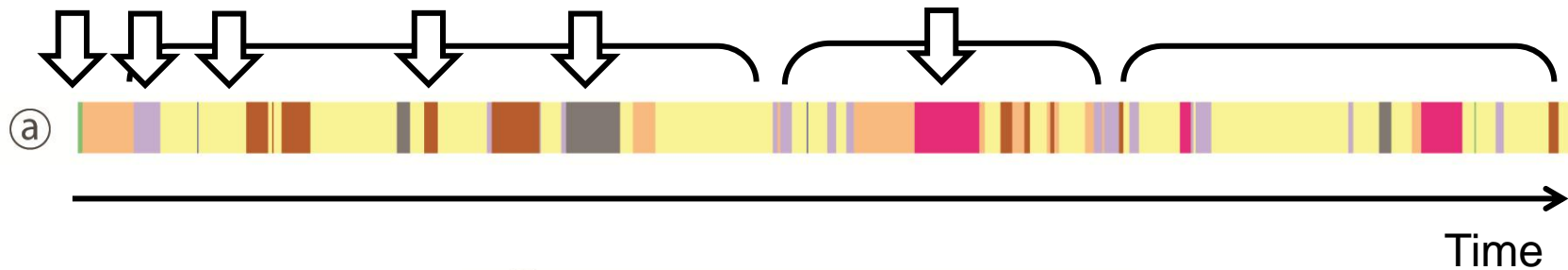
Fanny Chevalier

Christopher Collins

Ravin Balakrishnan

jianzhao@dgp.toronto.edu

http://www.cs.toronto.edu/~jianzhao

UNIVERSITY OF TORONTO

UOIT
CHALLENGE INNOVATE CONNECT

dgp
dynamic graphics project

RESULTS: PATTERNS OF USAGE

Time

Results: User Feedback

- "I used to draw trees by hand according to the output text files, so I could only compare 2 or 3 trees at a time and they are basically simple trees around 5 levels. With DAViewer I can compare many large trees efficiently. All trees are nicely aligned and the interactions allow me to focus on subtrees easily."

- "I know I want the tree levels aligned from the bottom, but I can't draw such layout from the output file in one round since the tree nodes are indented in the depth-first order."

- "I used to insert debug code inside the parsers, for a verbose output, but it is tedious and almost impossible to find patterns. Now, I can do it visually by just creating what I need and clicking a button."