DAViewer User Instructions Manual

Jian Zhao
Department of Computer Science, University of Toronto
jianzhao@dgp.toronto.edu

DAViewer is an interactive visualization system for assisting computational linguistics researchers to explore, compare, evaluate and annotate the results of discourse parsers. A copy of the code is on GitHub: https://github.com/jeffjianzhao/DAViewer. A journal publication [1] about this tool and the accompanying video demo can be found online at http://www.cs.toronto.edu/~jianzhao.

Interface Overview

The interface of DAViewer is shown in Figure 1 below. It is comprised of: (a) an overview which displays the overall performance of all the parsers over all the documents in the dataset, (b) a detail panel which visualizes the discourse tree structures of the focused algorithms and documents as node-link or icicle dendrograms (see [1]), (c) a status panel which provides the basic properties of the currently selected items as well as an interactive legend for filtering operations, (d) an annotation panel which allows users to edit notes, (e) a text panel which shows the content of the active document as parsed by the focused algorithm, and (f) a search window which for querying based on keyword and structure over the data in the detail panel.

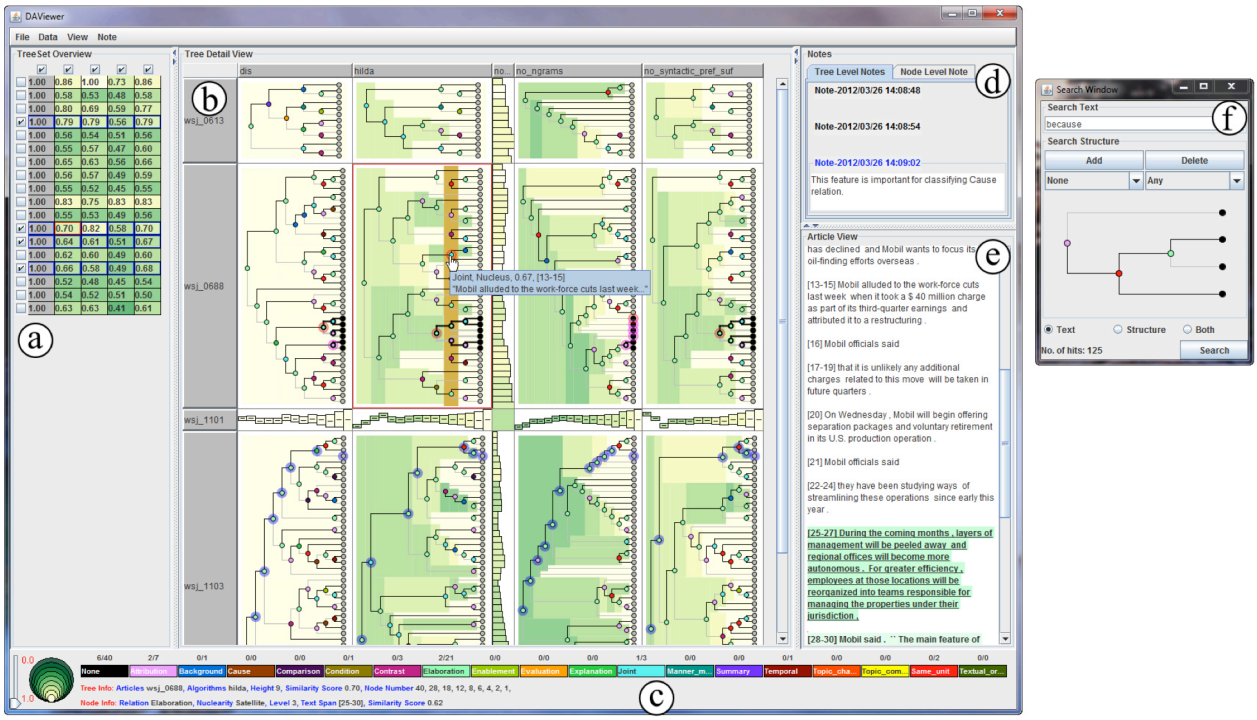

Figure 1 The DAViewer Interface.

Data and Overview Panel

DAViewer can load three kinds of datasets: 1) a single discourse tree in *.out.hilda.tree* format or *.out.dis* format; 2) a tree set (i.e., a matrix of discourse trees) in *plain text* format with markups of the metadata as shown in the sample dataset; and 3) a workspace of the system in *.tree.wksp* format which saves the interface status (e.g., annotations) along with the discourse trees.

Visualizations in Overview Panel

The overview panel (Figure 1a) consists of a matrix view representing the entire available dataset—a collection of discourse trees, each tree resulting from the computation of a given parsing algorithm (columns in the matrix) applied to a given document (rows in the matrix). The overview is useful to determine at a glance how the different algorithms perform on the same document and how the same algorithm performs across a set of documents as compared to the reference algorithm (i.e., the gold standard, or otherwise user-defined). One column is set as the reference and shown in grey. In Figure 1a, the leftmost column serves as the reference. Each non-reference cell is colour-coded according to a similarity score between the tree in that cell, and the current reference tree in that row (i.e., we compare trees for the same document). The numeric scores are displayed in each cell, and also mapped to a yellow-to-green scale. In our current implementation, the similarity measure relies on the element-based measure proposed by Bremm et al. [2].

File Menu and Data Menu

The menu items in the file menu are:

- Open workspace – open a dataset in .tree.wksp format.
- Save workspace – save the current application status into .tree.wksp format.
- Open treeset – load a new dataset of discourse trees into the system in the plain text format.
- Import tree(s) – import a single tree in .out.hilda.tree format or .out.dis format or a treeset in the plain text format into the selected cells in the current overview.
- Exit

The menu items in the data menu are:

- Set reference column – set the resulting discourse trees in the selected column as the reference column for computing the similarity scores.
- Add column – add an empty column to the overview matrix.
- Remove columns – remove selected columns in the overview matrix.
- Add row – add an empty row to the overview matrix.
- Remove rows – remove selected rows in the overview matrix.

Tree Detail View

In order to preserve a consistent tabular representation across the views and maintain the alignment of EDUs across algorithms, the selection is algorithm- and document-based, in that the user selects the set of trees of interest by checking row (document) and column (algorithm) headers in the overview. All the cells at the

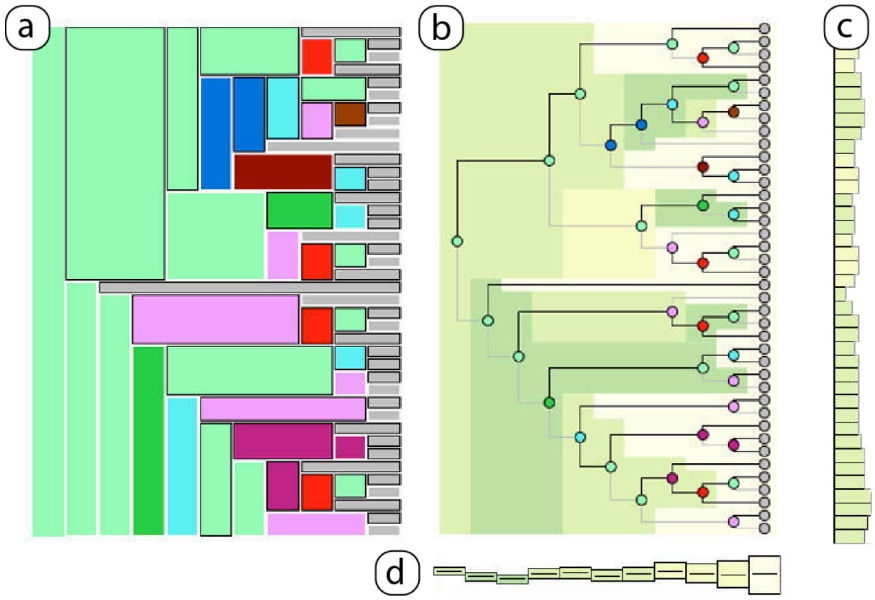

Figure 2 Discourse tree visualizations.

intersection of the selected rows and columns are loaded in the detail panel, as shown in Figure 1b. The currently selected cells are outlined in blue in the matrix overview.

Visualization and Interaction of Discourse Trees

We designed two main types of representations to visualize the discourse trees: an adapted version of the icicle plot (Figure 2a), and a dendrogram—a branching node-link diagram particularly suited to reflect relationships (Figure 2b), that resembles the traditional representation. In both representations, the nodes are colour-coded according to the assigned relation label from the 18 described by Hernault et al. [3]. The specific hues were selected from 18 of the 22 most distinguishable colours in Green-Armytage's colour alphabet [4]. The hue of the links between the nodes and their parent relation indicates the nuclearity of the children nodes in the node-link dendrogram: black indicates a nucleus (the most prominent nodes), and grey indicates a satellite. In the icicle-dendrogram, the colour of the outline of the children node indicates this property. To facilitate the comparison of trees and sub-trees, we build an icicle-dendrogram where nodes are colour-coded according to the similarity scores of the internal nodes using a yellow-to-green palette. This dendrogram is used as a background on top of which the node-link dendrogram is overlaid (Figure 2b), thus serving as a heatmap where errors are made more salient because of the darker colour.

We also propose compacted rows and columns, while providing a vertical (Figure 2c) or horizontal (Figure 2d) informative representation that summarizes important characteristics of the compacted discourse trees. In the vertical compact view, each bin corresponds to a leaf node, i.e., an EDU. The width of a bin encodes the depth of its associated leaf in the tree, i.e., the number of intermediate clusters the EDU belongs to on the path to the root.

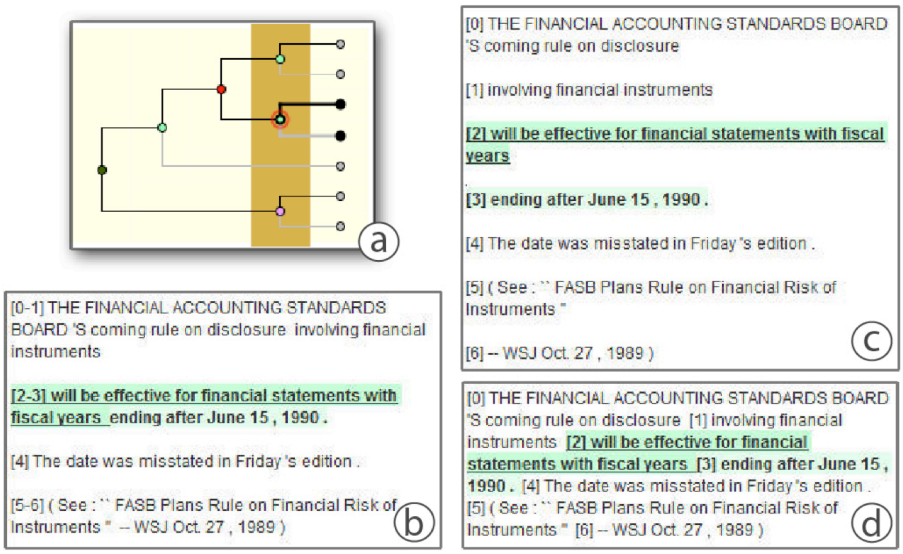

Figure 3 Article view.

The colour of a bin is mapped to the average similarity score of its corresponding leaf node and all nodes on the path from the leaf to the root. In the horizontal compact view, each bin corresponds to a level of the compacted tree. The height of a bin encodes the number of nodes at the associated level. The colour is mapped to the average similarity score of the set of nodes. We also use the vertical axis to represent the centroid vertical position of the nodes belonging to the level, and center the bins accordingly. A horizontal black mark in the center of the bin is used as a visual cue of the centroid location.

DAViewer provides fluid interactions for visual exploration of discourse trees. The user can double-click the row or column table header of the detail view to display discourse trees in the compact views. As the user moves the cursor in the active table cell of the detail view, a semi-transparent brown ribbon indicates the tree level where the cursor resides, as an aid to identify the nodes on the same tree level. When the user selects a node to access detailed information of the node and its subtree in the status panel (Figure 1c), it also triggers the emphasis of the node and the branch under it, using thicker edges. In other trees in the same row, the branches corresponding to the same EDUs (leaf nodes) are also emphasized (see middle row, Figure 1b) so as to facilitate the comparison of internal structures. DAViewer also supports traditional tree operations such as branch collapsing when the user double clicks on a node. It is also possible to collapse branches through batch operations using a context menu (e.g., collapse all branches under a node with similarity score below a certain threshold, or all branches up to a specific tree level).

Article View

When the user select a node in the tree detail view, the associated text of the EDUs under the selected node is loaded into the text panel (Figure 1e), and emphasized in several ways: (1) we apply a bold font to the selected node's content, (2) the text of the first child node is underlined, and (3) the background is coloured according to the main relation's hue, with a dark or light tone whether the text belongs to a nucleus or a satellite branch respectively (Figure 3).

Text Display Formats

There are three ways the text can be formatted. First, clicking on an empty space within a level (as opposed to a node) selects that level, which rearranges the text into a layout we call hybrid display: the text is split into paragraphs according to the EDUs branching at that level (Figure 3b). This text display is particularly useful when the user wants to analyze the intermediate stages of the relation grouping. The second text layout is the separated display (Figure 3c) which presents each EDU as a separate paragraph. Finally the continuous display (Figure 3d) displays the article as a single paragraph. The separated display is suited to the matching EDUs to leaves in the tree, as the EDUs remains clearly separated, while the continuous display aims to offer a more comfortable way of reading text in a continuous flow. The separated display is equivalent to choosing the leaf level in the hybrid display; the continuous display is equivalent to selecting the root level in the hybrid display.

View Menu

The menu items in the view menu are used to configure the visualizations across the matrix overview, tree detail view and article view.

- View selected trees – display the selected trees in the matrix overview in the tree detail view.
- Add selected rows/columns – add discourse trees in the selected rows or columns in the matrix overview into the tree detail view.
- Remove selected rows/columns – remove discourse trees in the selected rows or columns in the matrix overview from the tree detail view.
- Node-link view – change the tree view visualization in detail view to the dendrogram view.
- Icicle-plot view – change the tree view visualization in detail view to the dendrogram icicle-plot hybrid view.
- Hybrid text view – change the text display formats to hybrid text view.
- Continuous text view – change the text display formats to continuous text view
- Separated text view – change the text display formats to separated text view

Querying and Filtering

The status panel acts as an interactive legend for filtering operations, including the similarity score legend (Figure 4a), to filter out nodes with specific similarity scores, the relation legend (Figure 4b), where each label is a modal button for dynamic filtering.

DAViewer also incorporates querying functions for node and branch searching, including keyword search, structure-based querying, and the combination of the two (i.e., all branches that satisfy both queries) through a separate dialog (Figure 1f). Keywords can be entered in the text area; triggering the highlighting of all the nodes which descendant EDUs contain such a keyword. To perform a structural search, the user builds her pattern of interest either from scratch, or by editing a structure copied from the detailed view. The user can specify or omit the relation and nuclearity types associated to the nodes of the query pattern. When no value is specified, the engine searches for any value of such nodes. All the structures that match the query are highlighted in the detail panel with a blue halo. The structure-based query function has been explicitly requested by our users as a critical function to locate recurrent error patterns, as this task is extremely difficult to perform with their current analysis method.

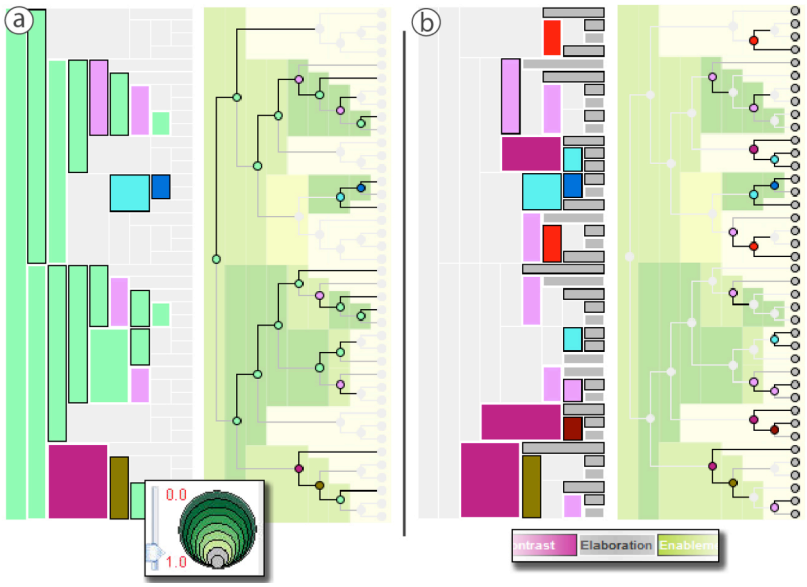

Figure 4 Filtering operations.

Annotations

DAViewer supports annotations on trees or nodes, and groups of trees or nodes through the panel as shown in Figure 1d. Notes are presented as a collapsible list of items titled with the timestamp of the last addition and can be saved together with the workspace format. The note panel is fully coordinated with the other views, implying that when a note is selected, all the corresponding items in the overview and detailed view are visually emphasized.

Note Menu

The menu items in the note menu are used for creating and managing the annotations associated with the tree(s) or node(s).

- Add notes to selected trees – create an empty annotation area on the "tree level notes" tab of the note panel for the selected tree(s).
- Add notes to selected nodes – create an empty annotation area on the "node level notes" tab of the note panel for the selected node(s).
- Remove selected note – delete the selected annotation from the note panel.

References

1. J. Zhao, F. Chevalier, C. Collins, and R. Balakrishnan. Facilitating Discourse Analysis with Interactive Visualization. In *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 18(12), pp. 2639–2648, 2012.
2. S. Bremm, T. von Landesberger, and M. Heb. Interactive visual comparison of multiple trees. In *Proc. of the IEEE Symp. on Visual Analytics Science and Technology*, pages 29–38, 2011.

3. H. Hernault, H. Prendinger, D. A. duVerle, and M. Ishizuka. Hilda: A discourse parser using support vector machine classification. In *Dialogue and Discourse*, 1(3):1–33, 2010.
4. P. Green-Armytage. A colour alphabet and the limits of colour coding. In *Colour: Design and Creativity*, 5(10):1–23, 2010.