WE1S "pyldavis" module
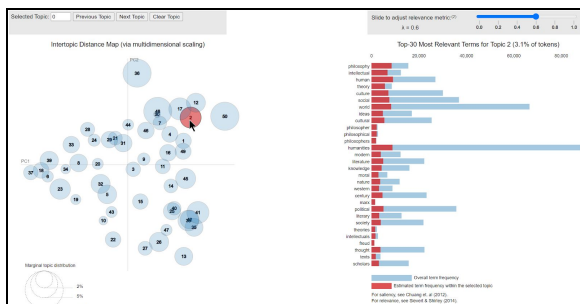
Included in the WE1S Workspace (see S-2, the "pyldavis" module contains a Jupyter notebook for creating a pyLDAvis visualization (one tool in the WE1S Topic Model Observatory).[1]

The parent of this tool was LDAvis, which Carson Sievert and Kenny Shirley created for the R programming environment (see their article). Ben Mabey ported LDAvis to Python as pyLDAvis[2]

pyLDAvis is a general-purpose topic model visualization interface useful for getting an overview of a model, looking closely at topics, and studying words associated with topics. Among general-purpose interfaces, it stands out especially for the "relevance metric" slider that allows the user to adjust the ranking of words in a topic for better understanding of what the topic is "about."

The pyLDAvis interface has a **left** and **right** panels as shown in the screenshot below (click for larger image):

Left panel: Shows the topics in a model represented as circles, where circle size indicates the relative statistical weight of topics. Clicking on a circle, or entering a topic number in the search field, selects a specific topic for examination. Topics are positioned and spaced in an "intertopic distance map" giving a sense of the statistical nearness or farness of topics.

Right panel: Shows top words associated with the topic selected in the left panel, along with bar graphs for the words' relative weight in the model. The blue bar for any word represents that word's frequency in the overall topic model. The red bar represents a word's frequency within the specific topic. A "relevance metric" slider scale at the top of the right panel controls how the words for a topic are sorted (according to the notion of the "lift" of a word in a model as defined in Sievert and Shirley's article). Hovering the mouse over a word changes the left panel to show only the other topics in which that word is prominent.

WE1S's chapter on pyLDAvis in its *Topic Model Observatory Guide* provides goal-directed instructions for using pyLDAvis..

Further information:

* Carson Sievert and Kenny Shirley, "LDAvis: A Method for Visualizing and Interpreting Topics" (2014)
* Jeri E. Wieringa, "Using pyLDAvis with Mallet" (2018)
* WE1S *Topic Model Observatory Guide*: pyLDAvis

Live example for a WE1S topic model: C-1.50 (50 topics)

Jupyter notebook in this module:
* `create_pyldavis.ipynb`

WE1S module code source: [TBD] (MIT License)

---

[1] To use this module, a topic model for a project in the WE1S Workspace must previously have been made with the "topic_modeling" module (see S-13). See M-2 for a general explanation of topic models.

[2] pyLDAvis was not designed to use MALLET topic modeling data out of the box. WE1S's pyLDAvis transforms the MALLET state file into the appropriate data formats. The code for this adapted from a blog post by Jeri E. Wieringa.