National Library of Medicine

# Semantic MEDLINE Quick Tour

Final Report

February 2014

Christian Minter

NLM Associate Fellow, 2013-2014

**Project Sponsors:**

Dr. Tom Rindflesch, Lister Hill Center for Biomedical Communications

Suzy Roy, MEDLARS Management Section

## Table of Contents

[Abstract 3](#_Toc381956488)

[Background 4](#_Toc381956489)

[Methods 4](#_Toc381956490)

[Results 6](#_Toc381956491)

[Discussion 7](#_Toc381956492)

[References 7](#_Toc381956493)

[Acknowledgements 7](#_Toc381956494)

[Appendix 1: Technical Specifications & Metadata 8](#_Toc381956495)

[Appendix 2: Quick Tour Script 9](#_Toc381956496)

## Abstract

**Objective:** The purpose of this project was to create a quick tour video that introduces Semantic MEDLINE (SemMed) to librarians and scientists, and provides a brief demonstration of how to use its features.

**Methods:** The specific content for the video was determined through conversation between the Associate and the project sponsors. A storyboard was created and reviewed by the project sponsors, and staff from the MMS Training and Outreach Unit. After two storyboard revisions, the content was approved. The quick tour was created using Camtasia software and sent to the project sponsors and MMS training staff for review. After their approval, the video will be presented to the SemMed team, and eventually published on the Semantic MEDLINE webpage with the documentation.

**Results:** An eight minute and thirty-one second video was created. The video explains the purpose of Semantic MEDLINE, defines a predication, and demonstrates the search, summarization, and visualization features. The video will be made available to the public online.

**Conclusions:** The quick tour will help in efforts to market Semantic MEDLINE. It will also provide users with information about semantic predications and Semantic MEDLINE’s features in an accessible format. Camtasia Studio was fairly easy to use, however for a first-time user, the video editing process was more time consuming than expected.

## Background

Semantic MEDLINE (SemMed) is under development at the Lister Hill Center for Biomedical Communications. It is a web application that manages the results of a PubMed search. Natural language processing is used to extract semantic predications from titles and abstracts. The predications represent the content of the article, and are presented in a graph that includes links to the original MEDLINE citations. SemMed is useful for keeping up with current research, as well as for scientific literature-based discovery.

In 2012, a former Associate Fellow explored the potential for SemMed’s use as a collaborative research tool for scientists and librarians. It was found that both scientists and librarians were eager to use SemMed, but were in need of a web-based tutorial that would explain how they could access and use the tool. At the moment, most users have found out about SemMed from published articles, conference presentations, or word of mouth. As part of the next step initiatives, the SemMed team would also like to market the tool to a wider audience. The quick tour will help with these publicity efforts.

## Methods

The project started with a meeting between the Associate and the project sponsors, Dr. Tom Rindflesch and Suzy Roy to discuss the project goals. The discussion covered background information, the purpose of the quick tour, the intended audience, and the intended use. The purpose of quick tour is to introduce SemMed and demonstrate its features. It will also define the phrase “semantic predication” as the SemMed team has encountered many users who confuse predication with prediction. The intended audience at the moment is librarians and scientists. The quick tour will be used to market SemMed to this user community, and remain available online as an instructional resource for users to access on their own.

For the screen capture software, the Associate compared Techsmith Camtasia Studio and Adobe Captivate. She decided to use Camtasia because reviews stated that it was easier for a first-time user to learn, and there appeared to be more NLM staff familiar with Camtasia than Captivate. The Associate already had Camtasia version 7 installed on her computer. However, she requested the purchase of a license to the updated version 8, as well as a high-quality microphone headset. Both requests were approved and processed without difficulty.

In preparing to create the quick tour content, the Associate reviewed published literature and past presentations on SemMed, reviewed other NLM quick tours and learning resources, and performed practice searches in SemMed. In preparing to create the video, she reviewed Lynda.com tutorials on Camtasia, and consulted with MMS librarian Kate Majewski to discuss 508 compliance, and tips for getting started with the software.

The Associate developed a storyboard using MS PowerPoint software. The purpose of the storyboard is to plan out the video’s content and design. The storyboard consisted of screenshots, actions on the screen (such as callouts or mouse movements), and the accompanying text to be used for audio narration and closed captioning (see Appendix 2 to read the script). The storyboard was reviewed by Dr. Rindflesch and Suzy Roy, as well as MMS staff Janet Zipser, Kate Majewski, and Steve Emrick. They reviewed the storyboard for content, design, communication style, and clarity in demonstrating and explaining SemMed. Two revisions were made to the storyboard.



**Figure 1. Sample storyboard PowerPoint Slide**

After the storyboard was approved, the Associate began recording the quick tour. The video was recorded in short clips, to make it easier to edit afterwards. The screen captures were recorded first, and then the audio narration. Callouts and zooming were added next. Then all the clips were joined into one video. The last step was to add the closed captioning. Dr. Rindflesch, Suzy, and Kate reviewed the video. No revisions were needed. The Associate will present the video to the SemMed team on February 28th. After the presentation, the Associate and the project sponsors will address any feedback from the SemMed team, and decide which webpage(s) would be the best location for posting the quick tour online.

****

**Figure 2. Camtasia Studio Interface (version 8.2)**

## Results

An eight minute and thirty-one second quick tour was completed. The quick tour was originally intended to be three to five minutes in length, but the time limit was extended in order to incorporate all the necessary content. The quick tour content includes:

* an introduction to SemMed,
* the definition of a semantic predication,
* a demonstration of the search, summarization, and visualization features,
* options for users to find more information about SemMed.

The web version of the video includes closed captioning to meet 508 compliance. It also includes a search feature that allows the user to search the text of the script and skip to the part of the video that contains a specific word or phrase.

## Discussion

The quick tour video will help the SemMed team to enhance their promotion efforts for Semantic MEDLINE. It provides a short and basic introduction, and demonstrates the most important aspects of the tool that a new user needs to know. The accessible format will allow users to access it remotely from their own computer, whenever and wherever they need it. This video is the beginning of a suite of SemMed learning resources. Additional quick tours are planned for the future.

There were small challenges in determining how much time and effort certain phases of the project would require. The Associate underestimated the amount of time that it would take to address storyboard revisions, and as a first-time user of Camtasia, the video retakes and editing were a little more time-consuming than expected.

Overall, this project has been very beneficial to the Associate, and enabled her to gain new knowledge and skills. In the process of creating the quick tour, she learned how to use Semantic MEDLINE, gained new software skills, and became more familiar with 508 compliance standards and instructional design concepts.

## References

Rindflesch, T. C., Kilicoglu, H., Fiszman, M., Rosemblat, G., & Shin, D. (2011). Semantic MEDLINE: An Advanced Information Management Application for Biomedicine. *Information Services & Use*, 31, 15-21.

Roy, S. (2012). *Librarian Assistance in the Use of Semantic MEDLINE: Outreach to Scientists.* (Unpublished spring report). National Library of Medicine, Bethesda, MD.

## Acknowledgements

The author would like to thank the project sponsors, Dr. Tom Rindflesch and Suzy Roy, as well as MMS staff Janet Zipser, Kate Majewski, and Steve Emrick. She is also thankful for the support of Dr. Kathel Dunn, the other Associate Fellows, her preceptor Janice Kelly, and NLM senior staff, Dr. Donald A.B. Lindberg, Betsy Humphreys, Joyce Backus, and Dianne Babski.

## Appendix 1: Technical Specifications & Metadata

**Recording**

Software: Techsmith Camtasia Studio v. 8.2; MS PowerPoint 2010 (for intro & ending slides);

Mozilla Firefox web browser

Hardware: Microphone headset – Hamilton Electronics model HA5USBSM

Recording area dimensions: 1000 x 752 (custom setting)

Logo(s) used: UMLS

**Production** (custom settings)

* Video dimensions: 800 x 600
* MP4 Flash/HTML5 player
* Features include: closed captioning, searching
* Controller: classic theme; after video stop with replay button; pause at start with automatic thumbnail

**Metadata**

Project Info:

* Title: Semantic MEDLINE Quick Tour
* Subject: Semantic MEDLINE
* Category: Training
* Language: English
* Description: A brief overview of Semantic MEDLINE, and a demonstration of its use.

Author Info:

* Creator: National Library of Medicine
* Publisher: National Library of Medicine
* Email: custserv@nlm.nih.gov
* URL: <http://skr3.nlm.nih.gov/SemMed>
* Rights management: Government information at NLM Web sites is in the public domain. Public domain information may be freely distributed and copied, but it is requested that in any subsequent use the National Library of Medicine (NLM) be given appropriate acknowledgement.

HTML Options:

Title: National Library of Medicine: Semantic MEDLINE Quick Tour

## Appendix 2: Quick Tour Script

(Title Slide)

Unified Medical Language System

Semantic MEDLINE: A Quick Tour

(Introduction to SemMed Slide)

Semantic MEDLINE is a web application that summarizes citations returned by a PubMed query. Natural language processing is used to extract semantic predications from titles and abstracts. The predications are presented in a graph that has links to the original citations. By displaying the meaning found in the literature, Semantic MEDLINE helps you to manage your query results and decide which articles to read.

(Predication Definition Slide)

A semantic predication represents a relationship asserted in text. It consists of two arguments (subject and object) joined by a relation (or predicate). For example, the predication “Tamoxifen TREATS Malignant neoplasm of Breast” has subject “Tamoxifen,” object “Malignant neoplasm of Breast,” and relation “TREATS.” The predications represent the content of the article and form the basis for further processing.

(Search Tab)

* We will start our quick tour at the Search tab.
* You can enter any PubMed search into the Query box. We will search for citations and predications linking Parkinson's disease and resveratrol.
* The Most Recent drop-down menu specifies the number of citations most recently added to PubMed that will be retrieved in the search. For our search, we will use the default setting, and retrieve only the 500 most recent citations.
* Use the Start Date and End Date to set the date range for the search. It’s best to keep the default setting for the end date, as it reflects the last load of predications to Semantic MEDLINE. We will keep the default setting for both dates.
* PubMed Limits provide most of the options that are available in PubMed Filters to help you narrow your search results. For our search, we will not use PubMed Limits.
* Click on Search.
* Our query on Parkinson’s disease and resveratrol resulted in 64 citations. The list of results displays the authors, title, and PubMed ID.
* The PubMed ID or PMID, is the unique identifier assigned to each article’s record in PubMed.
* Clicking on the authors’ names will take you to the citation in PubMed. Clicking on the Abstract link will display the text of the abstract.

(Summarization Tab)

* Click the Summarization tab.
* Summarization identifies key predications from the citations that are retrieved in the search. The summarizations are then visualized as a graph.
* 528 predications have been extracted from the citations retrieved from our search on Parkinson’s disease and resveratrol.
* There are three options to narrow the focus of the summary to your research interests. The first option is Summary Type. Let’s select treatment of disease, since our search for resveratrol focuses on the treatment of Parkinson's disease, and not the underlying biology of interactions with other substances.
* The second option is More Relations. If unchecked, it will omit predications that occur less frequently. Let’s check the box to include all predications.
* The third option is to Select a UMLS concept. These concepts are additional topic terms. They are listed in descending order based on their frequency of occurrence in the predications. The list of concepts will change based on the summary type selected. Let’s use the UMLS concept of Neurodegenerative Disorders.
* Click on Summarize. Let’s scroll down the page to view the summary of the predications.
* On the right side, we see the Subject, Predicate, and Object that make up the predication. In this example, the predication is “Estrogens PREVENTS Nerve Degeneration”.
* To the left of the predication, the sentence that it was extracted from is displayed.
* To the left of the predication sentence, the PMID is listed. Clicking on the PMID will take you to the citation in PubMed.
* Let’s scroll back to the top of the page.

 (Visualization Page)

* Click on Visualize.
* On the Visualization page, predications are displayed as a graph. Nodes represent the subject and object, and **arcs** represent the relation.
* Clicking on Relation Labelsdisplays a legend for the graph. Arcs are color-coded to represent the type of relation. Nodes are color-coded into semantic groups that represent biomedical categories. You can read the predications in the graph by looking at the nodes, and the color and direction of the arcs connecting the nodes. Here, we see that “Parkinson Disease ISA Neurodegenerative Disorders”.
* The checkboxes let you choose which relations to display on the graph. Uncheck TREATS. All the arcs and nodes for the predications about treatment disappear from the graph. Check the TREATS box, and the related predications display on the graph again.
* Clicking on a node displays Concept Informationat the top of the information panel. Let’s click on Neurodegenerative Disorders. This concept appears 14 times in the graph. Buttons display links to other National Library of Medicine resources that provide information on the concept. In this example, we have links to the UMLS Metathesaurus and Genetics Home Reference (GHR).
* Clicking on an arc displays Relationship Informationabout the predication at the bottom of the information panel. Let’s click on the arc between resveratrol and Neurodegenerative Disorders. The relationship information displays the subject (resveratrol), the relation (TREATS), and the object (Neurodegenerative Disorders).
* The Relationship Information also gives the frequency of occurrence of the predication and the number of citations in which the predication appears. This predication occurs seven times in the graph, and is supported by six citations. A predication can appear in more than one citation, and a citation can have more than one predication as well.
* Clicking on the Citations button opens a window showing the text that contains the predications. The specific sentence that the predication was extracted from is highlighted in orange. Scroll down to see all the citations in the list. The PMID, title and abstract of each citation are displayed. Clicking on the PMID will take you to PubMed. Click Close to return to the graph.
* Sometimes you may need to change the graph layout to make it easier to read. You can change the shape of the graph by using the Layout drop-down menu. Let’s change the shape to radial. Make the graph larger by clicking the magnifying glass with the plus sign, or make the graph smaller by clicking the magnifying glass with the minus sign. To move the graph, click on it and drag it to the new position on the screen. You can click and drag nodes too. Let’s move the node, resveratrol.

(More Help Slide)

For more help on using Semantic MEDLINE, click on the SemMed Documentation link. It is visible on both the Search and Summarization pages.

(Closing Slide)

* To start your search, go to the Semantic MEDLINE website. A free UMLS license is required for access and is available at the site.
* For questions or feedback, go to UMLS Customer Service.