NASA/TM–2018-220078

IBM Watson Supporting Space Radiation

Lisa A. Scott Carnell
Langley Research Center, Hampton, Virginia

Theodore D. Sidehamer
Craig Technologies, Hampton, Virginia

Wade A. Hunter and Jeremy J. Yagle
Langley Research Center, Hampton,Virginia

July 2018

NASA STI Program . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counter-part of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at http://www.sti.nasa.gov

- E-mail your question to help@sti.nasa.gov

- Phone the NASA STI Information Desk at 757-864-9658

- Write to:
  NASA STI Information Desk
  Mail Stop 148
  NASA Langley Research Center
  Hampton, VA 23681-2199

NASA/TM–2018-220078

IBM Watson Supporting Space Radiation

Lisa A. Scott Carnell
Langley Research Center, Hampton, Virginia

Theodore D. Sidehamer
Craig Technologies, Hampton, Virginia

Wade A. Hunter and Jeremy J. Yagle
Langley Research Center, Hampton,Virginia

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

July2018

Available from:

NASA STI Program / Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199
Fax: 757-864-6500

ABSTRACT
The NASA Human Research Program (HRP) Space Radiation (SR) Program Element has been working with IBM Watson Explorer (WEX) to create a tool that allows researchers to search the NASA SR-funded research corpus to help streamline research and maximize efficiency. The entire corpus of publications from research funded by the NASA SR Program Element has been ingested into WEX to allow for examination of: synergies across funded research areas, gaps in research, and collaboration opportunities. This information will be valuable to both scientists and managers as it will allow analysis related to specific scientific questions, inform key decisions and support cross validation of study results. NASA will also evaluate the potential to make WEX publicly available in order to facilitate proposal generation and to enhance collaborations within and across disciplines.

INTRODUCTION
In the research community, a significant amount of data and information is being generated at a rate that is not possible for humans to keep pace with. For example, Baylor College of Medicine was able to analyze 240,000 articles in a few weeks using advanced data mining tools that would take a human an estimated 70 years to complete if they were reading at a rate of 10 articles per day. [Spangler et al. 2014] And even that is ambitious. In another example, Johnson and Johnson data experts were able to identify genetic profiles that respond to drugs without adverse side effects in a matter of days. This type of effort would typically take more than forty years to accomplish without advanced data mining and analysis tools. [Jeelani, 2014] The NASA SR Element is similarly challenged with staying on top of the data and publications being generated from internally-funded investigators and the external research community. Analytical tools are needed to parse the information in a timely manner. In order to address this issue, a pilot study employing IBM Watson Explorer was designed to create a platform that would help streamline research gaps, support evidence updates, facilitate collaboration amongst primary investigators and keep up to date with research in real time. Employing advanced analytical tools provides the ability to see unobvious linkages and relationships that will help determine the most promising research paths forward.

WHAT IS WATSON EXPLORER?
Watson Explorer (WEX) is a text analytics and data mining tool that evolved from the world famous IBM Watson that played and won on Jeopardy. WEX provides the ability to search and analyze large volumes of unstructured information from multiple sources to quickly understand and deliver relevant insight. It employs natural language processing that understands the way humans speak and discerns information based on context within a document.

HOW IS WATSON EXPLORER DEVELOPED?
To tailor the WEX tool to support Space Radiation, a subject matter expert (SME) worked closely with a NASA data science team to train the WEX tool on the specific area of interest and use. The SME identified key words, taxonomies and dictionaries that were used to develop a facet structure, which is a classification scheme used to systematically organize knowledge by applying semantic categories. The facet structure serves as the primary search feature of the tool. The data science team ingested publications and metadata provided by the SME to create the knowledge base, and also extracted data from relevant NASA-centric sites and databases. There are currently over 1,000 Space Radiation documents and associated metadata from HRP-

funded research available in the WEX tool, along with more than 500,000 metadata files related to space radiation research from PubMed. The data science team also ingested the NASA Space Life and Physical Sciences Research Applications (SLPSRA) Taskbook that houses all NASA-funded research from the SLPSRA Division along with the NASA Human Research Roadmap. This allows scientists and managers to examine NASA specific data regarding funding, research areas of focus and gaps in these areas to name a few.

USER INTERFACE

A screen capture of the WEX interface is shown in Figure 1. The interface layout can be customized to individual preferences for viewing content. Users can view details such as facets, facet values, documents, time series and trends that will change in real-time with each new search. The "Facet Navigation" field (Figure 1) is one of the key methods used to perform a search. It is comprised of various approaches that allow the user to search by parts of speech, phrases, document clusters, and/or concepts that are machine-generated in conjunction with a taxonomy defined by the SR subject matter expert (SME). The "Facet Values" indicate the frequency the selected facet appears in the search while the "Time Series" section graphs the number of publications generated by the search indexed by year of publication. The "Trends View" shows the frequency of a chosen phrase or search over time. In the "Documents" section, the user can review the associated abstract and metadata, and open the PDF if linked to the field.

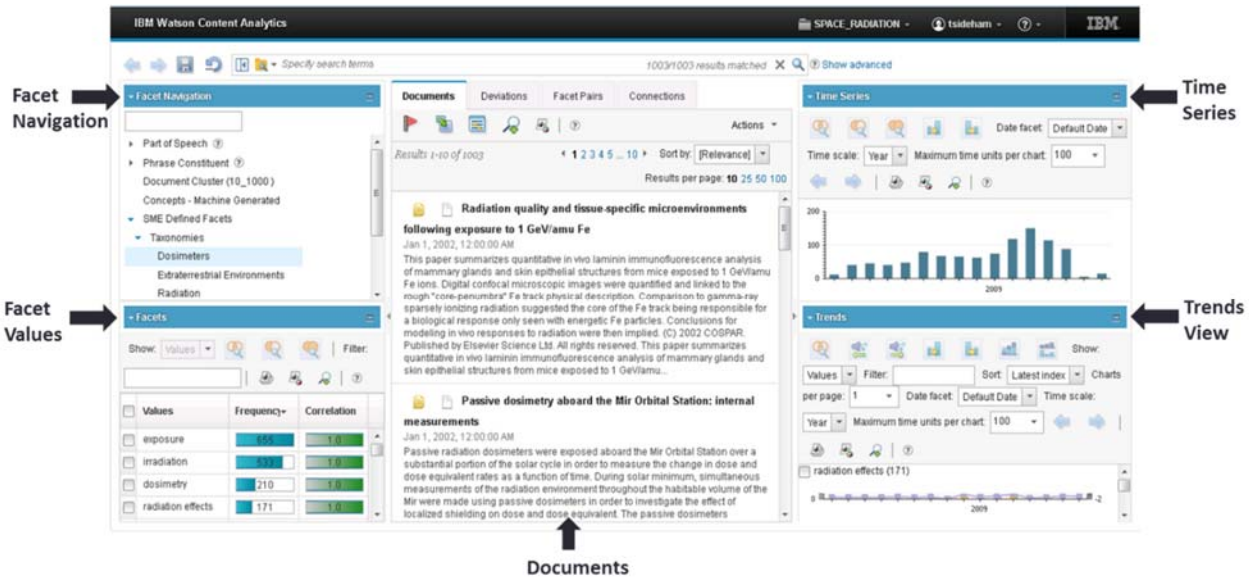

Figure 1. Illustration of screen shot for WEX. Site can be tailored to user preferences.

SEARCH FEATURES

So how do I search? The natural language processing feature is a powerful tool within WEX and can be used to take advantage of Watson Explorer's ability to parse text by phrases within the content selected. Figure 2 shows phrases or concepts that were machine-generated from

searching and parsing the documents within the HRP SR corpus. These phrases are the most common phrases found in the corpus of documents selected for this particular search and are listed in priority order by frequency (i.e., how many times those phrases appear). A user can search by multiple means and drill down into data within various searches. Facets help narrow down the search areas and the dual facet feature allows the user to combine multiple facets in a single search. Document searches can also be narrowed by selecting metadata, machine-generated concepts or user input under the "filter" or "specify search terms" sections. Each new search will yield an updated set of concepts and frequencies relative to those specific search criteria.

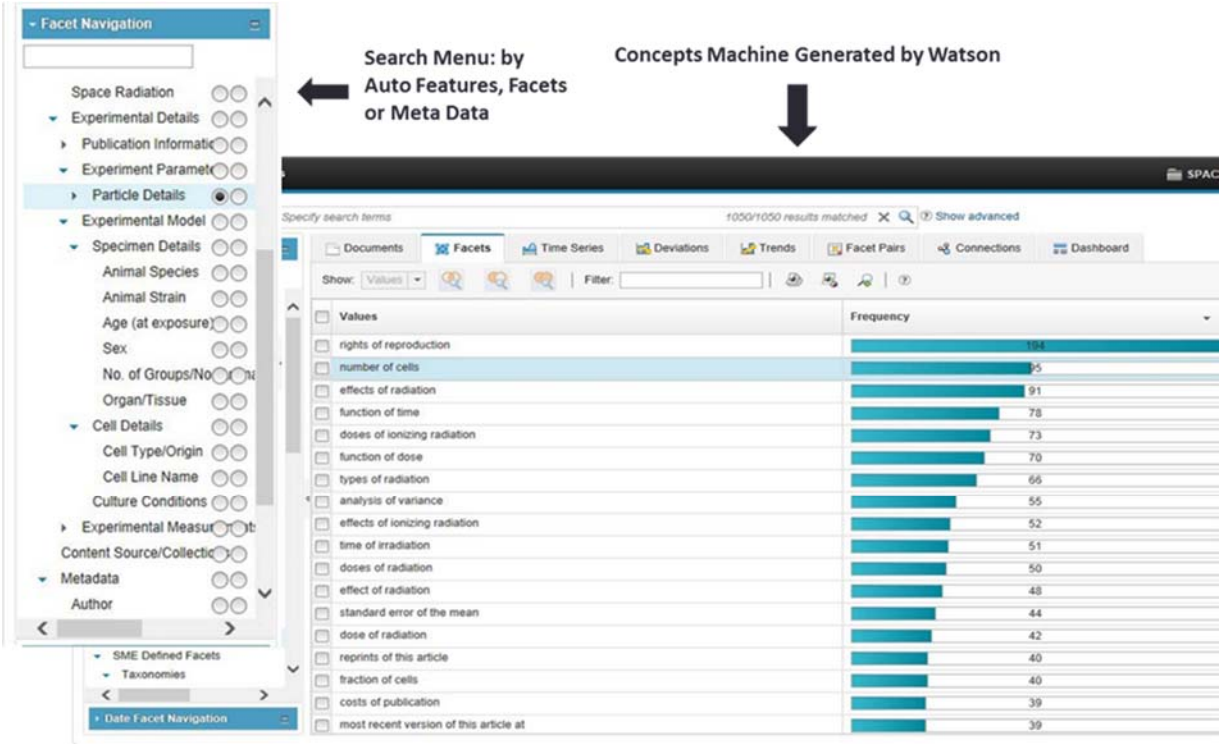

Figure 2. Demonstration of WEX ability to generate concepts for tailored searches.

EXPERT CONNECTIONS

Based on document authors, WEX can generate a network of connections within a corpus (such as HRP SR) or across a more narrowly selected set of documents that allows users to view linkages between or among those authors or "experts". For SR, this feature allows the user to explore working relationships between various investigators as well as connections between investigators and primary areas of research. The example in Figure 3 shows the key areas of research, connections (who the authors work with and publish with) and publications for a specific researcher. The top pane shows the entire connection network for a specific search with the red lines indicating a higher correlation between entities and yellow a lesser. The network view can be expanded by drilling in to examine linkages of interest as shown in the middle pane. Documents associated with the search network of interest are identified in the bottom pane of Figure 3.

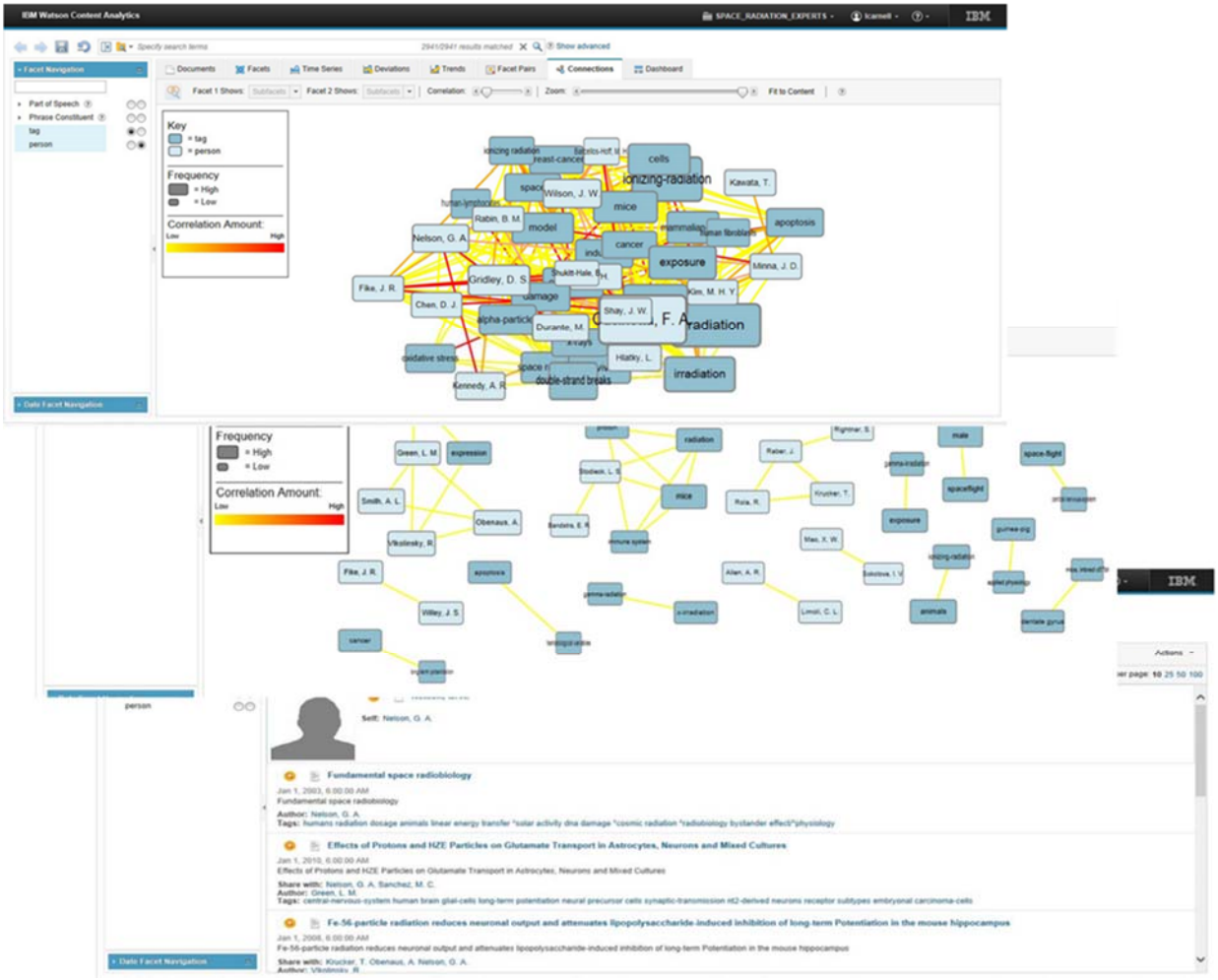

Figure 3. An example of an expert connection network using WEX.

KNOWLEDGE ASSISTANT

The WEX knowledge assistant can be valuable in finding answers to specific questions regarding a particular topic area. For example, Figure 4 illustrates the outcome when the user asks the question, "What animal models have been used in space radiation studies and more specifically, what strains, ages and sex?" Using the SR corpus and selecting "Animal Species" and "Animal Strain" facets, the knowledge assistant identified the mouse, rat, rabbit, ferret and minipig as the primary animal models employed, with the rodent models (rats and mice) being widely used in SR studies. The primary strains in HRP SR-funded studies for mice and rats have been heterozygous, ICR, CBA, nude, C57Bl/6J, and Sprague-Dawley. The Yucatan is the primary strain of minipig. The age range for the various species was between 5 weeks and 6 months, and both males and females have been studied similarly in rodents; however, rabbit studies have focused more on male models.

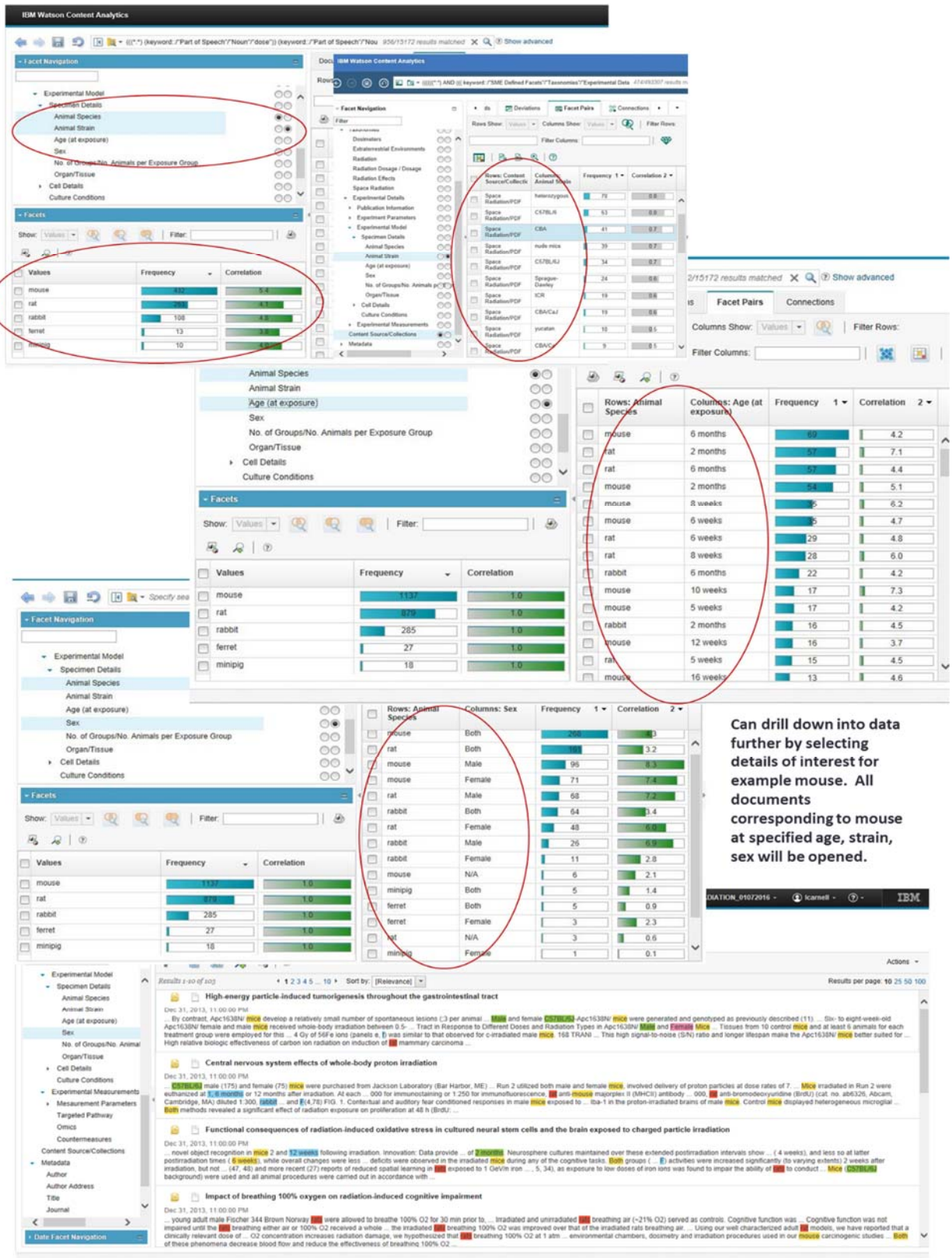

Figure 4. Demonstration of the breadth of information that can be obtained in minutes using WEX.

The knowledge assistant can further examine data for specific areas of interest. In the example below (Figure 5), the researcher is interested in the results on the brain of C57BL/6 mice exposed to radiation. By selecting C57BL/6, the knowledge assistant will show only those publications related to this study type and where this information can be found. The user can continue to refine the search, open the selections identified, create connection networks, generate time series or use a combination of these methods to answer the question of interest.

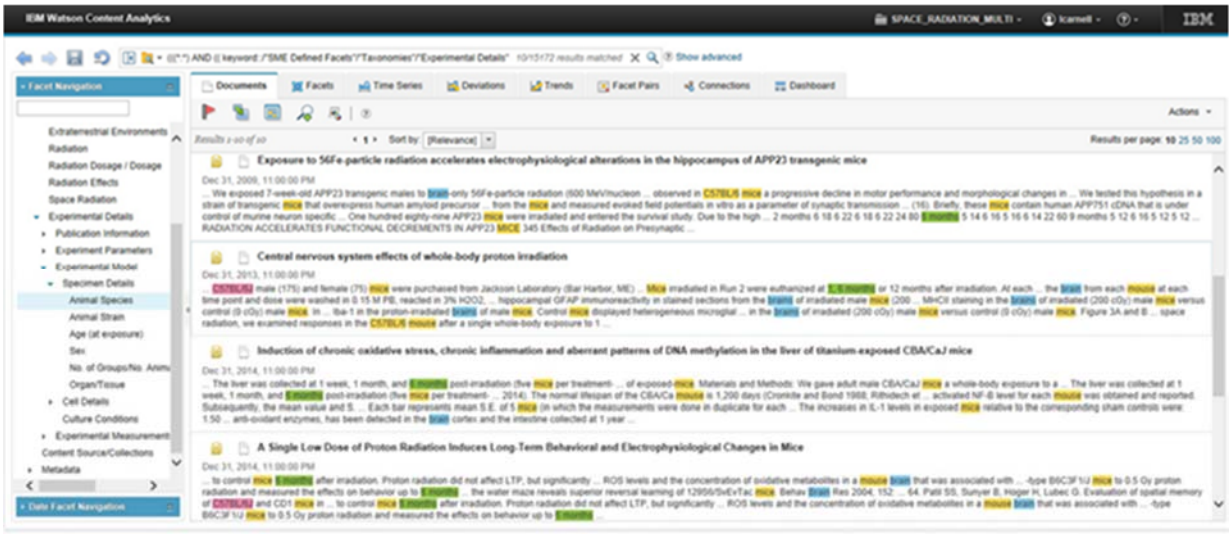

Figure 5. Example of a basic question for WEX and results obtained. The user can stop at this point and examine the data or continue to refine based on the specific question that is to be answered.

FUTURE WORK

Future goals include expanding the capability of this tool to incorporate documents and information from all elements of the Human Research Program and Space Life and Physical Sciences Research Applications Division, to leverage existing databases external to NASA, and to add advanced analytics to uncover relationships that were previously unknown for diseases arising from exposure to space radiation and other spaceflight stressors. WEX tools will help summarize the vast amounts of information and lead to refined risk assessment, biomarker discovery, and identification of biological countermeasures for mitigation and therapy. The next phase in the development of WEX for HRP and SLPSRA is to expand the existing capability by ingesting databases that will allow cross-referencing of relationships across multiple data sets and published literature. This advancement would provide deeper scientific understanding of critical space-related health problems and facilitate a data-driven process to achieve the goal of safely sending humans on extended deep space missions.

REFERENCES

Spangler S et al. Automated Hypothesis Generation Based on Mining Scientific Literature, KDD '14 Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. pp 1877-1886 New York, New York, USA — August 24 - 27, 2014

Jeelani M. WATSON Watson, Come here. I want you. Johnson & Johnson's CEO enlists IBM's big-data service to find new drugs. Fortune. 2014 Oct 27;170(6):36.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
1-07-2018

2. REPORT TYPE
Technical Memorandum

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE
IBM Watson Supporting Space Radiation

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Carnell Scott, Lisa A.; Sidehamer, Theodore D.; Hunter, Wade A.; Yagle, Jeremy J.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
651549.01.07.10

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
NASA Langley Research Center
Hampton, VA 23681-2199

8. PERFORMING ORGANIZATION REPORT NUMBER
L-20942

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSOR/MONITOR'S ACRONYM(S)
NASA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
NASA-TM-2018-220078

12. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified
Subject Category 62
Availability: NASA STI Program (757) 864-9658

13. SUPPLEMENTARY NOTES

14. ABSTRACT The NASA Human Research Program (HRP) Space Radiation (SR) Program Element has been working with IBM Watson Explorer (WEX) to create a tool that allows researchers to search the NASA SR-funded research corpus to help streamline research and maximize efficiency. The entire corpus of publications from research funded by the NASA SR Program Element has been ingested into WEX to allow for examination of: synergies across funded research areas, gaps in research, and collaboration opportunities. This information will be valuable to both scientists and managers as it will allow analysis related to specific scientific questions, inform key decisions and support cross validation of study results. NASA will also evaluate the potential to make WEX publicly available in order to facilitate proposal generation and to enhance collaborations within and across disciplines.

15. SUBJECT TERMS
HRP; SLPSRA; WEX; Watson; Analytics; Data mining; Space radiation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			STI Help Desk (email: help@sti.nasa.gov)
U	U	U	UU	12	19b. TELEPHONE NUMBER *(Include area code)* (757) 864-9658