

OAK RIDGE NATIONAL LABORATORY

MANAGED BY UT-BATTELLE FOR THE DEPARTMENT OF ENERGY

W. R. Besancenez
P.O. Box 2008
Oak Ridge, TN 37831-6163
Ph: (865) 576-1538, Fax: (865) 241-4082
INTERNET: besancenezwr@ornl.gov

June 17, 2011

To: All Interested Parties

Subject: Request for Information – Scientific Literature in Russian language

INTRODUCTION

UT-Battelle LLC is the managing and operating contractor of Oak Ridge National Laboratory (ORNL) under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). UT-Battelle LLC has recently established a system at ORNL called Foresight and Understanding from Scientific Exposition Network (FUSEnet) for its sponsor to enable computational research on automated techniques for text analysis. The primary focus is to develop automated methods that aid in the systematic, continuous, and comprehensive assessment of technical emergence using information found in the published scientific, technical, and patent literature. Knowledge and innovation are global phenomena and hence it is expected that novel findings and innovations are published across multiple languages. Generating new computational methods requires access to a large and diverse corpus of scientific and technical data, with data sets containing documents in the original language of publication.

The FUSEnet system is comprised of a protected, unclassified computer cloud operating on government computers at the ORNL Advanced Computing Center. The FUSEnet system is specifically designed to meet the computational scale and complexity requirements for large-scale data analysis and processing. A critical component of FUSEnet is the Document Repository (DR), a corpus of full-text scientific, technical, and patent documents in multiple languages. A user of the DR can only gain access through the security controls of FUSEnet. This is necessary to ensure the adherence to data licensing restrictions.

FUSEnet is a component of a larger program called Foresight and Understanding from Scientific Exposition (FUSE), under the sponsorship of the Intelligence Advanced Research Projects Activity (IARPA). IARPA is a US Government research sponsoring organization that invests in high-risk / high-payoff research programs. Public information about the FUSE Program can be found here: http://www.iarpa.gov/solicitations_fuse.html

LANGUAGE

The non-English language of interest to the FUSE Program under this RFI is Russian. Additional languages are of also of interest and have been discussed in separate RFIs.

CONTENT AND FORMAT

FUSE is interested in acquiring peer-reviewed scientific literature in Russian, including research articles, reference articles, and conference proceedings. Years of interest are 1980 – present.

FUSE seeks to acquire a minimum of 100,000 documents in the Russian language; a corpus containing tens of millions of documents in the Russian language is preferable.

FUSE is interested in acquiring content across all areas of science and technology research, including:

- Agricultural Science
- Biology, Biotechnology & Microbiology
- Biochemistry, Genetics, Immunology & Molecular Biology
- Chemistry
- Computer Science
- Ecology & Environmental Sciences
- Engineering
- Geosciences
- Materials Science
- Mathematics
- Neuroscience
- Physics
- Plant & Animal Science
- Space Sciences

FUSE requires data in a digital format, such as pdf or XML. A structured, machine-readable format, such as XML, is preferable. Respondents should characterize their available content, including counts of documents available by year, file format, and discipline. The options for regular data updates, such as, weekly, monthly, quarterly, twice a year, or annually, for example, should be discussed.

SOURCES

FUSE is interested in acquiring data from any vendor with the content licensing rights to provide agreeable terms and conditions, to include primary publishers and public or private data aggregators. Respondents should identify the data sources available in their collections.

LICENSING TERMS AND CONDITIONS

At a minimum, any license agreement shall allow for UT-Battelle LLC personnel and contractors, US Government staff personnel, US Government contractors, as well as independent contractors and academics working on behalf of the US Government to access and manipulate the data. The primary Users will be the UT-Battelle LLC staff, government employees and contractors directly supporting the FUSE Program. Additional Users will be those government employees, contractors, and grantees supporting affiliated efforts.

Terms and conditions shall allow data to be locally loaded on access-controlled government networks to include a program document repository at Oak Ridge National Laboratory. Secondary sites with similar or enhanced data protection capabilities may also be required to satisfy research needs. FUSE

will minimally need access to the data through July 2016, either through a perpetual license agreement, fixed term, or renewable options. A perpetual license is preferred, but not required.

Data is being acquired for research on large scale computer-based algorithms (called, for example, 'text analytics', 'text mining').

REQUEST FOR INFORMATION (RFI) RESPONSES

Responses are due by 4:00 p.m. local time on **July 15, 2011**. Responses shall be submitted by email to the undersigned.

Email: besancenezwr@ornl.gov

Responses to the RFI shall include the following information:

1. Description of the format that the data is available in
2. Characterize the available content, including counts of documents available by year, file format, and discipline.
3. Identify the data sources available in the collections.
4. Identify the cycle that regular data updates are available, (weekly, monthly, quarterly, etc.).
5. Identify the type of license agreement(s) that are available and provide examples.
6. Confirm that the Licensing Terms and Conditions stated in this RFI are acceptable.
7. Provide budgetary pricing for 100,000 documents in the Chinese language and up. Please provide pricing for all of the documents you have available in Chinese.

DISCLAIMER

Responses to this notice are not offers and cannot be accepted by UT-Battelle to form a binding contract. Respondents are solely responsible for all expenses associated with responding to this RFI. For programmatic reasons, this RFI has been issued in advance of receipt of funds that would permit a subcontract(s) to be awarded. Funds may become available within the next 12 months. No legal liability on the part of UT-Battelle for any payment may arise until funds are made available and a subcontract is awarded.

If there are any questions please contact me.

Sincerely,



William R. Besancenez

Email: besancenezwr@ornl.gov

Team Lead, Contracts Division,

Contractor to the US Department of Energy