NLP/Text Mining Post-doctoral Fellow – National Institutes of Health, Rehabilitation Medicine Department

Job title: Post-doctoral fellow  
Institution: National Institutes of Health (NIH)  
Department: Rehabilitation Medicine Department (RMD), Epidemiology and Biostatistics Section  
Location: Bethesda, MD (primary location) and one of three sites in Falls Church, VA; Baltimore, MD; or Washington, DC  
Duration: Multi-year (pending satisfactory annual review of progress)  
Who may apply: U.S. Citizens, permanent residents, and foreign applicants (eligible for a J1 visa) who obtained a PhD from an accredited university in the past 5 years

Summary of position: We are seeking a post-doctoral fellow to engage in a multi-year project using statistical natural language processing or text mining to analyze medical records and programmatic documents for the nation’s largest disability programs, Social Security Disability Insurance (DI) and Supplemental Security Income (SSI), administered by the Social Security Administration (SSA). This is an unprecedented opportunity to harness unstructured information from millions of documents, use this information to improve SSA’s disability adjudication process, and advance the body of medical knowledge related to disability. Our interdisciplinary program is designed to provide experiences that foster the development of skills necessary to conduct independent research and to cross-train post-doctoral fellows in disciplines that are not traditionally represented in Rehabilitation Medicine such as mathematics, statistics, economics, engineering, or computer science. Our training program is intended for students interested in population-level health services research and disability-related public health issues; however, our fellows are encouraged to become actively engaged in the rich and vibrant NIH research community. This is a fully funded multi-year post-doctoral fellowship.

Research areas: natural language processing, text mining, machine learning

In the Epidemiology and Biostatistics Section of the RMD we hire people with a broad range of analytical and technical skills who are ready to tackle great challenges and make an impact on the lives of people with disabilities. We are currently looking for an NLP/Text Mining researcher who can develop techniques for extracting and understanding information from medical records and other text documents.

Role: As a postdoctoral fellow, you will conduct research in NLP/Text Mining, develop novel algorithms, implement prototypes, present and publish your work. You will research and identify new technologies, systems or applications of relevance to SSA’s and NIH’s NLP vision. You will collaborate with a multi-disciplinary team of scientists on a wide range of problems. This position will bring analytical rigor and statistical methods to the challenges of improving the accuracy, efficiency, and timeliness of SSA’s disability programs.

Responsibilities:

• Research, develop, and apply methods to tackle some of the challenges faced by SSA’s disability programs by tapping into information from medical records and other text documents
• Implement prototypes
• Conduct analyses on large datasets
• Develop new algorithms and methods for optimizing the disability determination process
Minimum qualifications:

- PhD in computer science, computational linguistics or related field from an accredited university
- Dissertation in natural language processing/text mining, or machine learning
- A minimum of 18 semester hours in computer science, computational linguistics, statistics, mathematical statistics, or probability theory
- Excellent quantitative and applied research skills
- Strong oral and written communication skills
- Ability to work effectively in a multidisciplinary team setting
- Demonstrated proficiency in programming with a high-level programming language (e.g., Java, Python, C++, C#)
- Knowledge of state-of-the-art NLP/Text Mining methods

Preferred qualifications:

- Experience with several of the following NLP subdomains: named entity recognition, semantic and grammatical disambiguation, negation and uncertainty identification, relationship extraction, temporal relationship extraction, information extraction, spelling error identification and recovery, ontological reasoning, text mining
- Experience with the following NLP tasks: tokenization, part of speech tagging, morphological decomposition, chunking, segmentation, regular expressions
- Experience with machine learning algorithms and optimization techniques
- Experience applying deep learning techniques to text and image data
- Experience with Optical Character Recognition (OCR) and Optical Mark Recognition (OMR)
- Experience working with large bodies of text
- Experience using NLTK
- Considerable practical experience in quantitative analysis
- Proficiency programming in a statistical language (e.g., Stata, SAS, R, MATLAB, GAUSS, Mathematica)
- A proven high level of organizational and communication skills
- Demonstrated ability to work effectively in a multidisciplinary team setting
- Demonstrated ability to provide training at various levels
- Demonstrated experience preparing conference presentations, research findings, and manuscripts for submission to peer-reviewed journals
- Ability to conduct independent research and to form internal and external collaborations in support of research initiatives in conjunction with a multi-disciplinary team in a fast-paced environment
- Knowledge of and adherence to the ethical guidelines and regulatory requirements for research involving human subjects
- A strong interest in topics related to health and disability, related programs and policies, and/or health services.

Application instructions:
For prompt consideration email the documents listed below to Daniel Hobbs, Management Analyst, Epidemiology and Biostatistics Section, National Institutes of Health, Mark O. Hatfield Clinical Research Center, Rehabilitation Medicine Department, Email: Daniel.Hobbs@nih.gov. Please refer to this announcement as the “NIH-RMD SSA NLP/Text Mining post-doctoral fellowship_[LAST NAME]”.

Application materials:
• curriculum vitae
• a summary of past research experience
• graduate transcript
• a copy of one of your published papers
• a list of at least three references