

AI Scientist (Semantic Search & Natural Language Processing)

Roivant Sciences New York, NY

About the job

Roivant Sciences' mission is to systematically reduce the time, cost, and risk of delivering new medicines to market by accelerating the development of promising drug candidates to improve the lives of patients and their families. The company is privately held, has in-licensed 25+ assets, and is very well-capitalized. The Computational Research team is focused on building an analytical engine to enable the identification and valuation of drugs for potential acquisition.

Title: AI Scientist (Semantic Search & Natural Language Processing)

Location: New York, NY

Manager: Director, Computational Research

Position Summary: We are looking for an AI Scientist with experience in Natural Language Understanding applied to information extraction to help build and optimize the unstructured data ingestion pipeline that underlies our pharmaceutical analysis engine. This pipeline enables us to aggregate information about drugs across various dimensions. The data sources are diverse in origin and massive in scope, requiring classification, ranking, information extraction and linkage to enable drug identification and monitoring of the scientific progress made in disease treatment.

Semantic Search Powering

Key Responsibilities:

- Build the tools necessary to optimize the use of our internal document semantic search engine
- Understand the architecture of the open source document indexing (Apache SOLR) and search engine Open Semantic Search (OSS)
- Implement the use of ontologies/dictionaries, exploring OSS features in depth to fully leverage those tools (enable tagging documents with relevant entities)

- Improve and update ontologies/dictionaries
- Develop processes to update and debug OSS when new releases or patches are available
- Optimize our unstructured data collection pipeline

Information Extraction And Linkage

- Implement state-of-the-art Named Entity Recognition for multiple entities leveraging Multi-task learning
- Establish relationship extraction between entities with classification of relationship types
- Perform topic modeling on documents in combination with sentiment analysis

Insight Extraction

- Leverage the information extracted from documents to design algorithms to alert analysts of important catalyst events
- Design and implement the storage of the extracted information in a graph database
- Use knowledge graph inference to propose un-witnessed relationships between entities (establish new links) and support data linkage corrections (remove wrong links)

Skills & Qualifications

- Proficient in Python with experience in understanding existing complex code
- Experience working in Linux environment and using GitHub
- Experience reproducing published results and improving on them using Tensorflow
- Experience developing machine learning models at scale from inception to business impact
- Successful experience in collaborating on code development
- Ability to partner with team members using strong written and verbal communication
- Experience in graph database design, query language and knowledge graph algorithms (such as neo4j)
- Familiarity with cloud-hosted distributed computing (AWS EC2)
- Familiarity with knowledge graph inference and interfacing with Neural Network preferred
- Experience in active learning, web scraping, and/or developing and executing a research agenda are ideal
- Self-motivated, proactive, and able to thrive in ambiguous situations

Experience & Education Requirements

- PhD and 2+ years' experience (preferred) or Master's degree and 5+ years' work experience
- Experience working in the pharmaceutical industry is a plus

- Authorized to work in the United States

Working Conditions

- Dynamic, interactive, fast-paced office environment
- Limited domestic or international travel may be required

Compensation

- Competitive compensation and benefits package, including initial equity options grant
- Equity options granted annually for top performers

Seniority Level

Associate

Industry

- Computer Software
- Financial Services
- Pharmaceuticals

Employment Type

Full-time

Job Functions

- Other