ExploreTerra

Unearth Your Energy Potential

Training

NATURAL LANGUAGE PROCESSING

Course Start Date: April 11, 2024
Registration open

Embark on a transformative journey with our NLP course, meticulously crafted to rapidly advance your comprehension and mastery of Natural Language Processing

ExploreTerra

Unearth Your Energy Potential

WELCOME!

Ph.D. Justo taught Natural Language Processing (NLP) at McGill University. Drawing on his extensive industry experience, he has optimized the course curriculum to provide students with practical, real-world knowledge and skills in this rapidly evolving field

SCOPE

Course Start Date: April 11, 2024
Registration open

- **Duration:** 3 months (flexible, based on interest and demand).
- **Hours:** 28 hours. 4 hours of consulting and assistance.
- **Price per person:** $840 CADs, $620 USD
- **Target Audience:** Anyone with basic Python knowledge.
- **Format:** Theory and practice, with focus on the particular needs of the student.
- **Focus Areas:** Machine learning, Data analysis, Python programming.
- **Type or training:** Remote (Teams) or in person.
- **Maximum number of students:** 20
- **Software:** Python, Anaconda, Jupyter Notebook, or Similar
- **Languages:** Available in English
- **Certification:** A certification will be issued upon completion of the training.

ExploreTerra
Unearth Your Energy Potential

CONTENT

Week 1: NLP Software Setup and Linguistic Foundations

- Python libraries for NLP (NLTK, spaCy, Hugging Face's transformers)
- Input/output of NLP systems (text formats, character encodings)
- Basic text processing (regex, basic tokenization)
- Overview of linguistic concepts (morphology, syntax, semantics)

Week 2: Data Pre-processing Techniques in NLP

- Advanced tokenization techniques (sentence splitting, subword tokenization)
- Text normalization challenges (Unicode normalization, transliteration)
- Handling multilingual and domain-specific corpora
- Building custom normalization functions

Week 3: Methods to Contextual Embeddings

- Advanced vectorization techniques (Hashing Vectorizer, Word Embeddings)
- Introduction to distributed representations (skip-gram, CBOW)
- Overview of contextual word embeddings (BERT embeddings)
- Text vectorization for downstream tasks (classification, clustering)

CONTENT

Week 4: Supervised, Unsupervised, and Semi-supervised Learning

- Deep diving into classification and clustering algorithms for text data
- Ensemble methods and hyperparameter tuning
- Semi-supervised learning techniques (self-training, co-training)
- Evaluation strategies (holdout, K-fold cross-validation, stratification)

Week 5: From RNNs to Transformer Architectures

- Detailed architecture of RNNs, LSTMs, GRUs - benefits and drawbacks
- Introduction to Attention Mechanisms and the Transformer Architecture
- Detailed explanation of pre-trained models (BERT, GPT-2, T5)
- Fine-tuning best practices and common errors

Week 6: Cutting-edge NLP and ethical consideration

- Natural Language Understanding (NLU)
- Natural Language Generation (NLG)
- Discussion of challenges in deploying NLP systems (performance, scalability, maintenance)
- Vector databases in production (recommendation systems, search)
- Ethics in NLP (data privacy, model biases, fairness)

CONTENT

Week 7: Maximizing Efficiency and Quality in Machine Translation

- Overview of low-resource languages and their characteristics
- Data collection and annotation strategies for low-resource contexts
- Active Learning
- Transfer learning and cross-lingual embeddings
- PEFT: LoRa

Week 8: Beyond Positive and Negative - Complex Sentiment and Emotion Detection

- Sentiment Analysis with Deep Learning: Aspect-based sentiment analysis
- Emotion detection: Recognizing a range of emotions from text
- Lexicon-based approaches vs. machine learning approaches
- Advanced techniques: Sarcasm detection, negation handling, and irony

Week 9: Making NLP Models Understandable to Humans

- Explanation types: Model-agnostic vs. model-specific interpretations
- Tools for interpreting machine learning models (LIME, SHAP)
- Techniques for visualizing and explaining transformer-based models
- Importance of interpretability for trust and ethics in NLP

CONTENT

Week 10: Building Intelligent Agents That Can Communicate Naturally

- Overview of conversational AI: Task-oriented and open-domain dialog systems
- Structure of a dialog system: NLU, dialogue management, NLG
- Rule-based vs. machine learning approaches in dialogue systems
- Recent advances: Conversational models with memory, personalization

Week 11: Complex NLP Systems

- Reinforcement Learning with Human Feedback (RLHF)
- Similarity search in NLP
- Retrieval-Augmented Generation (RAG)
- Active Learning and Annotation strategies (uncertainty sampling, diversity sampling)
- Introduction to vector databases (FAISS, Annoy)

ExploreTerra

Unearth Your Energy Potential

ABOUT **COMPANY**

ExploreTerra's vision is to contribute to the geoscience consultancy and training landscape. Our core purpose is to establish a dynamic platform that creates connections between available talent and opportunities or needs within the energy industry.

We are dedicated to enriching the energy sector through specialized services and empowering geoscientists with technical training, tailored technology transfer, and the adoption of integrated, multidisciplinary best practices.

ExploreTerra
Unearth Your Energy Potential

YOUR INSTRUCTORS

Justo Rodriguez, PhD

Machine Learning Expert

Machine Learning Engineer, with a PhD in Chemical Physics, and half a decade of expansive experience, contributed to critical projects with renowned clients such as AAMI, Allianz, AstraZeneca, and Quest Diagnostics.

Recognized for leadership in a groundbreaking project in the CGI Global Innovation Challenge, along with a remarkable authorship of 30+ peer-reviewed scholarly articles. Expert in leveraging large language models (LLMs), natural language processing (NLP), and resource-efficient microservices to devise data-centric solutions that have substantially reduced costs and optimized processes in numerous industrial sectors.

GET
IN TOUCH

Our training programs, designed by seasoned geoscientists, are tailored to meet contemporary industry needs.

We prioritize technical and core skill enhancement and the incorporation of advanced technologies, equipping professionals for the dynamic field of geoscience.

CONTACT US :

justo.rodriguez@exploreterra.net

www.exploreterra.net

contact@exploreterra.net

explore.terra.energy

ExploreTerra

ExploreTerra

ExploreTerra
Unearth Your Energy Potential