Training Course on Named Entity Recognition (NER) and Information Extraction

Professional Development Training Course Outline

Category	Data Science	**Duration**	10 Days
Base Fee	$3,000 USD	**Delivery Mode**	Online / On-site Hybrid

Course Introduction

Training Course on Named Entity Recognition (NER) and Information Extraction dives deep into **Named Entity Recognition (NER)** and **Information Extraction (IE)**, equipping participants with cutting-edge **Natural Language Processing (NLP)** techniques to transform unstructured text into valuable, **structured data**. In today's data-driven world, the ability to automatically identify and extract key information from vast amounts of text is paramount for **business intelligence**, **knowledge management**, and **automated decision-making**. This program moves beyond foundational concepts, focusing on **advanced machine learning** and **deep learning architectures**, including **Transformer models**, to achieve state-of-the-art accuracy in identifying and categorizing entities such as persons, organizations, locations, dates, and domain-specific terms.

Participants will gain hands-on experience with practical applications, learning to implement robust NER and IE solutions for diverse real-world challenges. From **automating data entry** and **enhancing search capabilities** to powering **sentiment analysis** and **fraud detection**, this course provides the expertise needed to unlock the hidden insights within textual data. Through a blend of theoretical understanding and practical exercises, attendees will master the art of building, evaluating, and deploying high-performance information extraction systems, leveraging the latest advancements in **AI for text analysis**.

Programme Curriculum

Training Course on Named Entity Recognition (NER) and Information Extraction

Introduction

Training Course on Named Entity Recognition (NER) and Information Extraction dives deep into **Named Entity Recognition (NER)** and **Information Extraction (IE)**, equipping participants with cutting-edge **Natural Language Processing (NLP)** techniques to transform unstructured text into valuable, **structured data**. In today's data-driven world, the ability to automatically identify and

extract key information from vast amounts of text is paramount for **business intelligence**, **knowledge management**, and **automated decision-making**. This program moves beyond foundational concepts, focusing on **advanced machine learning** and **deep learning architectures**, including **Transformer models**, to achieve state-of-the-art accuracy in identifying and categorizing entities such as persons, organizations, locations, dates, and domain-specific terms.

Participants will gain hands-on experience with practical applications, learning to implement robust NER and IE solutions for diverse real-world challenges. From **automating data entry** and **enhancing search capabilities** to powering **sentiment analysis** and **fraud detection**, this course provides the expertise needed to unlock the hidden insights within textual data. Through a blend of theoretical understanding and practical exercises, attendees will master the art of building, evaluating, and deploying high-performance information extraction systems, leveraging the latest advancements in **AI for text analysis**.

Course Duration

10 days

Course Objectives

Upon completion of this training, participants will be able to:

1. Grasp the fundamental principles of Named Entity Recognition (NER) and Information Extraction (IE) within the broader scope of **Natural Language Processing (NLP)**.
2. Implement and optimize advanced NER techniques including **rule-based systems**, **machine learning models (CRFs, SVMs)**, and **deep learning architectures (LSTMs, Transformers)**.
3. Identify and engineer relevant features for effective NER and IE model training, including **word embeddings** and **contextual representations**.
4. Build and train custom NER models for specific **domain-specific entities** and industries.
5. Extract meaningful relationships between identified entities, enabling the construction of **knowledge graphs**.
6. Identify and extract information about events, including their participants, time, and location from unstructured text.
7. Efficiently process and transform large volumes of **unstructured and semi-structured text** into actionable, structured data.
8. Critically assess the performance of NER and IE models using appropriate metrics like **precision, recall, and F1-score**.
9. Work proficiently with industry-standard NLP libraries such as **SpaCy, NLTK, and Hugging Face Transformers** for NER and IE tasks.
10. Develop strategies for handling **entity ambiguity** and performing **entity disambiguation** for enhanced accuracy.
11. Understand how NER and IE fit into larger **data processing pipelines** and integrate extracted information into databases and applications.
12. Leverage **generative AI models (LLMs)** and **prompt engineering** for advanced information extraction scenarios.
13. Understand the ethical implications and best practices for responsible development and deployment of NER and IE systems.

Organizational Benefits

- Transform vast amounts of unstructured text into **actionable insights**, improving **data-driven decision-making**.

- Significantly reduce manual effort and time spent on **data entry** and **information gathering**.
- Gain deeper understanding of market trends, customer sentiment, and competitive landscapes through automated text analysis.
- Automate processes like document classification, contract analysis, and customer support ticket routing.
- Build comprehensive **knowledge bases** and **ontologies** by structuring disparate textual information.
- Leverage cutting-edge **AI and NLP capabilities** to stay ahead in a data-intensive environment.
- Minimize the need for manual data processing, leading to significant cost savings.
- Improve the reliability and standardization of extracted data compared to manual methods.

Target Audience

1. Data Scientists and Machine Learning
2. NLP Researchers and Developers
3. Data Analysts and Business Intelligence Professionals.
4. Software Engineers.
5. Researchers and Academics.
6. Product Managers and Project Leads
7. Anyone with a strong programming background (preferably Python) and an interest in applying AI to text.
8. Information Architects and Knowledge Managers

Course Outline

Module 1: Introduction to Named Entity Recognition & Information Extraction

- Defining NER and IE: What they are and their importance in modern data ecosystems.
- Overview of NLP pipeline and where NER/IE fit in.
- Challenges in extracting information from unstructured text (ambiguity, context).
- Real-world applications across various industries (finance, healthcare, legal, media).
- **Case Study:** Identifying key entities in financial news articles to track company mentions and market sentiment.

Module 2: Traditional Approaches to NER: Rule-Based Systems

- Regular Expressions (Regex) for pattern matching and simple entity extraction.
- Building dictionaries and gazetteers for known entities.
- Developing rule-based systems: advantages, limitations, and maintenance.
- Introduction to finite-state automata and transducers for text processing.
- **Case Study:** Extracting dates and monetary values from scanned invoices using rule-based patterns.

Module 3: Machine Learning for NER: Feature Engineering

- Transition from rule-based to statistical methods.
- Introduction to sequence labeling tasks (BIO, IOB tagging schemes).
- Feature engineering for NER: word features, character features, Part-of-Speech (POS) tags, lexical features.
- Contextual features: n-grams, surrounding words.
- **Case Study:** Designing features for a machine learning model to identify product names in e-commerce reviews.

Module 4: Machine Learning for NER: Conditional Random Fields (CRFs) & Support Vector Machines (SVMs)

- Understanding the architecture and principles of Conditional Random Fields (CRFs) for sequence labeling.
- Implementing NER with CRFs using libraries like sklearn-crfsuite.
- Introduction to Support Vector Machines (SVMs) and their application in text classification.
- Comparing traditional ML approaches: strengths and weaknesses.
- **Case Study:** Building a CRF model to extract medical conditions and treatments from clinical notes.

Module 5: Introduction to Deep Learning for NER

- Limitations of traditional ML for complex text data.
- Fundamentals of Neural Networks: perceptrons, activation functions.
- Word Embeddings: Word2Vec, GloVe, FastText – capturing semantic relationships.
- Recurrent Neural Networks (RNNs): basic concepts and limitations.
- **Case Study:** Utilizing pre-trained word embeddings to improve NER performance for less frequent entities.

Module 6: Advanced Deep Learning Architectures: LSTMs & Bi-LSTMs

- Addressing vanishing/exploding gradients in RNNs: Long Short-Term Memory (LSTM) networks.
- Bidirectional LSTMs (Bi-LSTMs) for capturing both past and future context.
- Combining LSTMs with CRFs for enhanced sequence tagging.
- Training deep learning models for NER: data preparation, hyperparameter tuning.
- **Case Study:** Developing a Bi-LSTM-CRF model for identifying legal entities in contractual documents.

Module 7: The Rise of Transformers for NER

- Introduction to the Transformer architecture: self-attention mechanism.
- Pre-trained Transformer models: BERT, RoBERTa, XLNet, ELECTRA.
- Fine-tuning Transformer models for downstream NER tasks.
- Hugging Face Transformers library: practical implementation and best practices.
- **Case Study:** Applying a fine-tuned BERT model to recognize company names and stock symbols from financial reports.

Module 8: Information Extraction: Relation Extraction

- Defining relation extraction: identifying semantic relationships between entities.
- Rule-based and pattern-based relation extraction.
- Supervised learning for relation extraction: feature-based methods.
- Deep learning approaches for relation extraction (e.g., using BERT for sentence pair classification).
- **Case Study:** Extracting "Acquired By" relationships between companies from news articles to build an M&A knowledge graph.

Module 9: Information Extraction: Event Extraction

- Understanding events and their components (triggers, arguments, roles).
- Rule-based methods for event extraction.
- Machine learning and deep learning for event detection and argument role labeling.

- Challenges in event extraction: temporal reasoning, causality.
- **Case Study:** Identifying "Product Launch" events from press releases and extracting product names, dates, and associated companies.

Module 10: Advanced Topics in NER & IE: Entity Disambiguation & Linking

- The challenge of entity ambiguity (e.g., "Apple" as a company vs. fruit).
- Techniques for entity disambiguation: contextual clues, knowledge bases.
- Entity linking: connecting extracted entities to external knowledge bases (e.g., Wikipedia, Wikidata).
- Building and leveraging custom knowledge bases for entity linking.
- **Case Study:** Disambiguating person names in biographical texts and linking them to a comprehensive public figure database.

Module 11: Data Annotation & Dataset Creation for NER & IE

- Importance of high-quality annotated data for supervised learning.
- Annotation guidelines and best practices.
- Tools and platforms for manual data annotation (e.g., Prodigy, Label Studio).
- Active learning and semi-supervised learning for reducing annotation effort.
- **Case Study:** Setting up an annotation project to create a dataset for extracting specific legal terms from contracts.

Module 12: Evaluation Metrics and Model Debugging

- Standard evaluation metrics for NER: Precision, Recall, F1-score, Exact Match.
- Evaluating relation and event extraction models.
- Error analysis and debugging techniques for NER and IE models.
- Strategies for improving model performance: data augmentation, curriculum learning.
- **Case Study:** Analyzing the performance of a NER model on a challenging dataset and identifying common error patterns to refine the model.

Module 13: Integrating NER & IE into Real-World Applications

- Deployment strategies for NLP models (APIs, batch processing).
- Scalability considerations for large-scale information extraction.
- Integrating NER/IE outputs into databases, dashboards, and downstream applications.
- Pipelines for continuous information extraction.
- **Case Study:** Building a system to automatically extract key information from customer feedback emails and route them to relevant departments.

Module 14: Ethical Considerations and Bias in NER & IE

- Bias in training data and its impact on NER/IE model fairness.
- Mitigating bias in data collection and model training.
- Privacy concerns and data anonymization in information extraction.
- Ethical implications of automated decision-making based on extracted data.
- **Case Study:** Discussing potential biases in a NER model trained on news articles and strategies to ensure fair and equitable entity recognition.

Module 15: Future Trends in NER & IE & Practical Project

- Latest advancements: few-shot and zero-shot NER, prompt-based IE.
- The role of Large Language Models (LLMs) in advanced information extraction.
- Industry trends and emerging applications of NER/IE.

- **Practical Project:** Participants will work on a real-world NER/IE problem, from data collection and model training to evaluation and deployment, applying all learned concepts.
- **Case Study:** Exploring the application of a new LLM for extracting highly specific, nuanced information from research papers.

Training Methodology

This training course employs a blended learning approach, combining theoretical instruction with extensive hands-on practice.

- **Interactive Lectures:** Engaging presentations covering core concepts, advanced algorithms, and practical considerations.
- **Live Coding Demonstrations:** Step-by-step demonstrations of implementing NER and IE models using Python and leading NLP libraries.
- **Hands-on Labs & Exercises:** Practical coding exercises and mini-projects to reinforce learning and build practical skills.
- **Case Studies & Discussions:** Analysis of real-world scenarios and challenges, fostering critical thinking and problem-solving.
- **Group Activities:** Collaborative tasks to encourage peer learning and diverse perspectives.
- **Q&A Sessions:** Dedicated time for participants to ask questions and receive expert guidance.
- **Project-Based Learning:** A culminating project where participants apply their knowledge to a complete NER/IE solution.
- **Mentorship and Feedback:** Opportunities for personalized guidance and constructive feedback on coding assignments and project work.

Register as a group from 3 participants for a Discount

Send us an email: info@fineskilltrainingcenter.org or call *+254769199797*

Certification

Upon successful completion of this training, participants will be issued with a globally- recognized certificate.

Tailor-Made Course

We also offer tailor-made courses based on your needs.

Key Notes

Upcoming Scheduled Sessions

Start Date	End Date	Location	Price
23 Nov 2026	04 Dec 2026	Online	$2,000
23 Nov 2026	04 Dec 2026	Physical	$3,000

Start Date	End Date	Location	Price
23 Nov 2026	04 Dec 2026	Physical	$6,000
23 Nov 2026	04 Dec 2026	Physical	$6,000
23 Nov 2026	04 Dec 2026	Physical	$12,000
23 Nov 2026	04 Dec 2026	Physical	$9,000
23 Nov 2026	04 Dec 2026	Physical	$12,000
23 Nov 2026	04 Dec 2026	Physical	$12,000

Ready to enroll or request a customized program? Book online or contact us:

Register Online Now

Email: info@fineskilltrainingcenter.com | Website: www.fineskilltrainingcenter.com