Yogesh Haribhau Kulkarni • You
AI Advisor (Helping people/organizations in their AI j…
now • 🌐

While "Data is the new oil", it's often not directly usable in its raw form. We need to extract the required components from it. For textual data, this process is called Information Extraction (IE). 💻 📄

I've recently published a technical blog at Medium titled "Journey of Named Entity Recognition" 📝

In this, I cover the following subtopics:
🔍 What is Information Extraction (IE)? Its components, applications, the process to convert unstructured text to structured data, and dive into the critical component called Named Entity Recognition (NER).
🏷️ What is NER, and how does it work? See it in spaCy!
🛠️ Custom NER and IOB data format: The key to training domain-specific NER models.
🚀 A real-world use case, tackled through multiple approaches:
➡️ Rule-based regex examples with Python's re library
➡️ Machine Learning with CRF library
➡️ Deep Learning with LSTM in PyTorch
➡️ Generative AI few-shot learning examples
💡 Conclusions: While the approaches have become easier to use, debugging has become more challenging. A hybrid approach may be necessary depending on data, expertise, budget, and the required level of debuggability.

The blog post is packed with code snippets of examples on these topics. 💻 It is at https://lnkd.in/d2KWGCNh 📚

#InformationExtraction #NLP #NER #MachineLearning #DeepLearning #GenerativeAI #TechnicalBlogging #CodeSnippets #Python #PyTorch #CRF Explosion PyTorch Meta Facebook scikit-learn Ines Montani Matthew Honnibal #mvp #mvpbuzz #gde DBpedia Association Tom Soru

Journey of Named Entity Recognition

medium.com • 7 min read