

Information extraction via open LLMs



 Nexusflow



Haystack
by deepset





Idea

 **Extract information:**
for example, from an article about animals



1 Create an information extraction **Pipeline**
with  Haystack and  Nexus Raven

2 Define a **fake function** with parameters
for the desired information
*"habitat" \Rightarrow type: list of strings; description: list of
places where the animal lives*

3 Choose the **source**
an article about capybaras

4 Run the Pipeline
passing function definition + source

5 Get a parsed "function call", with the
extracted data
habitat=["Panama", "Colombia", ...]





1. Information extraction Pipeline



LinkContentFetcher

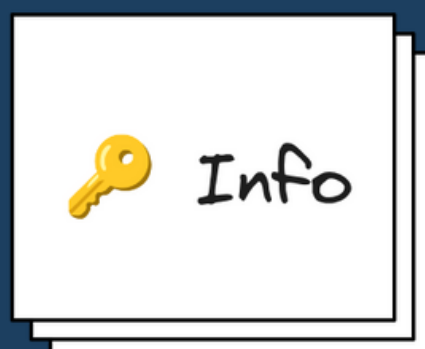
HTMLToDocument

DocumentCleaner

PromptBuilder

🧠 Generator/LLM
(Nexus Raven V2)

FunctionCallParser



2. Define a fake function with parameters for the desired information


```
def save_data (people: List[string],  
companies: List[string], summary: string,  
topics: List[string], about_animals:  
bool, about_ai: bool):  
    """  
    Save data extracted from source text  
  
    Args:  
    people (List[string]): List of the  
mentioned people  
    companies (List[string]): List of the  
mentioned companies.  
    summary (string): Summarize briefly what  
happened in one sentence of max 15 words.  
    topics (List[string]): what are the five  
most important topics?  
    about_animals (bool): Is the article  
about animals?  
    about_ai (bool): Is the article about  
artificial intelligence?  
    """
```



3. Choose the source

ARTIFICIAL INTELLIGENCE / MICROSOFT / BUSINESS

Sam Altman to return as CEO of OpenAI



After an attempted coup by OpenAI's board that lasted five days, Altman is returning alongside co-founder Greg Brockman.

By Nilay Patel and Alex Heath
Nov 22, 2023, 7:03 AM GMT+1

202 Comments (202 New)

Sam Altman will return as CEO of OpenAI, overcoming an attempted boardroom coup that sent the company into chaos over the past several days. Former president Greg Brockman, who quit in protest of Altman's firing, will return as well.


The company said in a statement late Tuesday that it has an "agreement in principle" for Altman to return alongside a new board composed of Bret Taylor, Larry Summers, and Adam D'Angelo. D'Angelo is a holdover from the previous board that initially fired Altman on Friday. He remains on this new board to give the previous board some representation, we're told.


4. Run the Pipeline


passing function definition + source




5. Get a parsed “function call” with the extracted data

 **people:** ['Sam Altman', 'Greg Brockman', 'Bret Taylor', 'Larry Summers', 'Adam D'Angelo', 'Ilya Sutskever', 'Emmett Shear']

 **companies:** ['OpenAI', 'Microsoft', 'Thrive Capital']

 **summary:** Sam Altman will return as CEO of OpenAI, overcoming an attempted boardroom coup that sent the company into chaos over the past several days.

 **topics:** ['OpenAI', 'Artificial intelligence', 'Machine learning', 'Computer vision', 'Natural language processing']

 **about_animals:** False

 **about_ai:** True