Sangrahaka

Annotation and Querying Tool

Hrishikesh Terdalkar, Arnab Bhattacharya
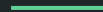
Department of Computer Science and Engineering
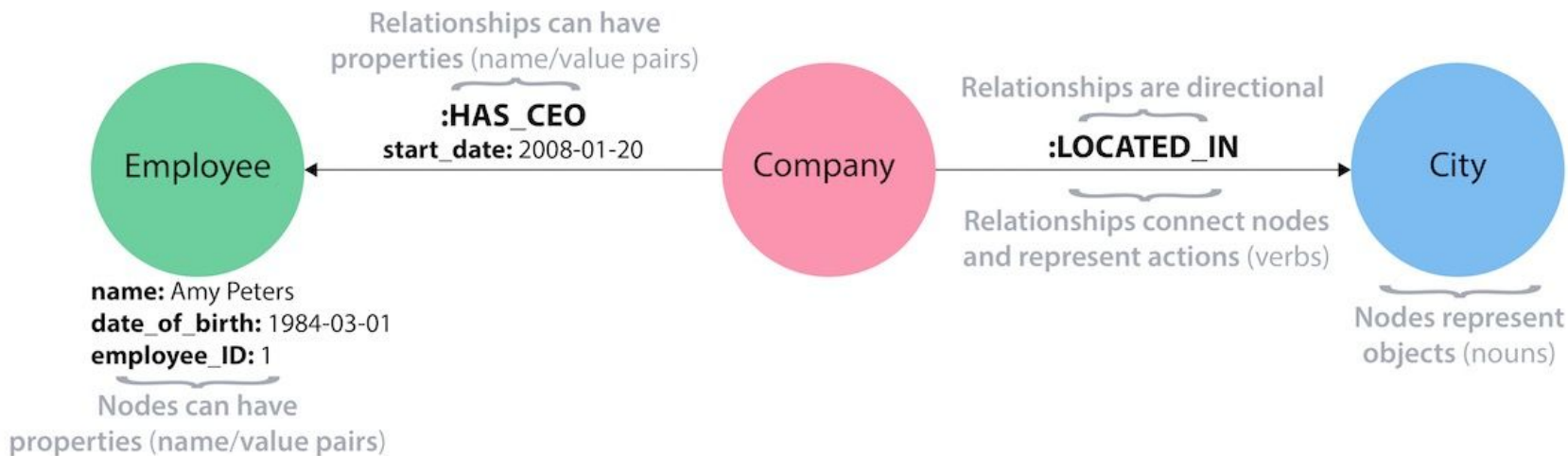Indian Institute of Technology Kanpur

Knowledge Graph

*.. a **knowledge base** that uses a **graph-structured** data model or topology to integrate data.* [Wikipedia]

Applications

- **Question-Answering**
- Semantic Search
- Other IR and NLP Applications

Knowledge Graph Example

Relationships can have
properties (name/value pairs)

:HAS_CEO
start_date: 2008-01-20

Employee

name: Amy Peters
date_of_birth: 1984-03-01
employee_ID: 1

Nodes can have
properties (name/value pairs)

Company

Relationships are directional

:LOCATED_IN

Relationships connect nodes
and represent actions (verbs)

City

Nodes represent
objects (nouns)

Property Graph Data Model

Source: https://neo4j.com/developer/graph-database/

Annotation

Process of marking relevant information from a corpus

Ontology-driven Annotation

- Entities as nodes
- Relationships as edges

Question Answering

Answering questions posed by humans

Querying Knowledge Graph

- Graph Database
  - Graph Query Language
- Query Templates
  - Natural Language Questions
  - Graph Queries
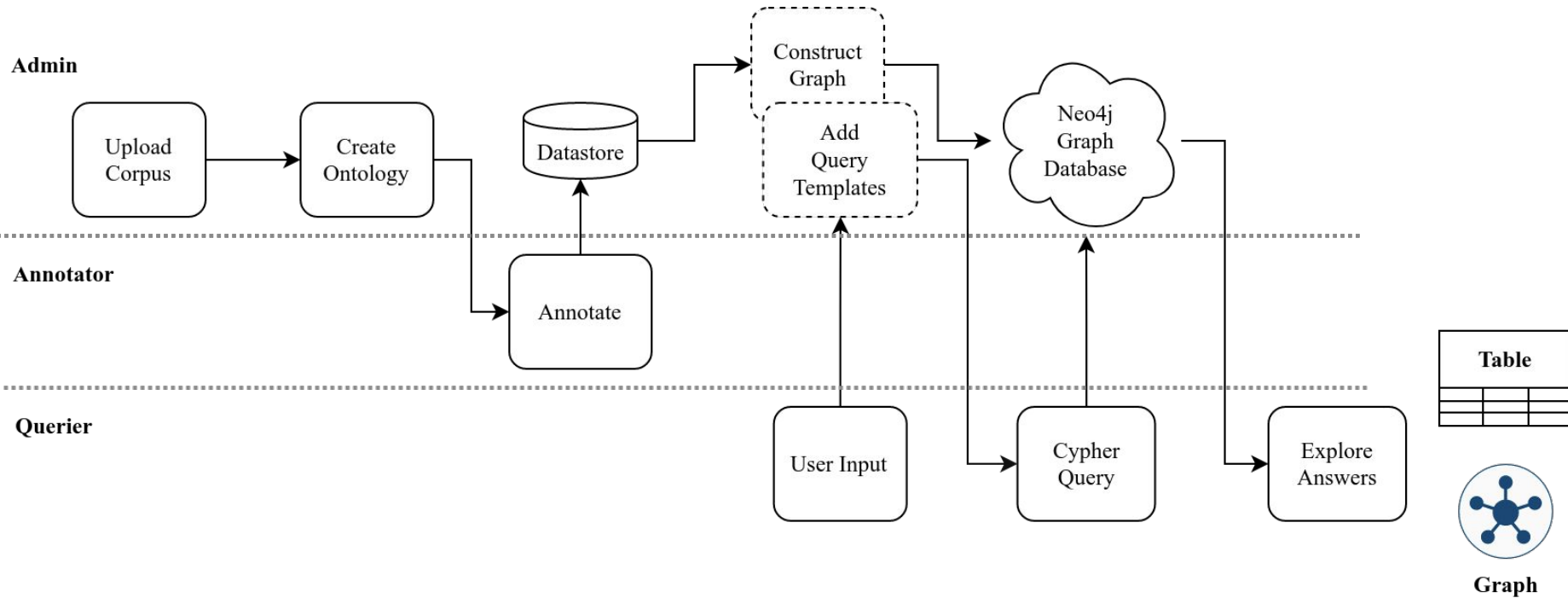
Sangrahaka

Annotation and **Querying** Tool

Focus

- Knowledge Graphs

Features

- User-friendly
- Web-based
- Configurable
- Customizable
- Fault-tolerant

Workflow

Features and Capabilities

Annotation

- Mark Nodes
- Mark Relations
- Adaptive Suggestions
- Multiple Annotators
- Multiple Curators

Querying

- Query Template System
- Interactive Query
- Cypher Query
- Results Explorer
  - Graph
  - Table

Administration

- Access Management
- Corpus Upload
- Ontology Creation
- Annotation Download

Configuration

- Single File Configuration
  - SQL, SMTP, ...
- Query Templates

Customization

- Structured Code
- Modular
- Examples

Other

- Easy Setup
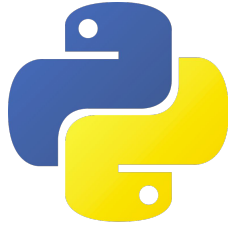- Web Deployment
- Fault Tolerance

Application

Real-world Annotation Tasks

Corpus	Lines	Annotators	Ontology		Annotations	
			Node Types	Relation Types	Entities	Relationships
BPN [1]	180	5	25	30	602	778
VR [2]	17655	12	114	144	3050	3989

[1] Bhavaprakasha Nighantu -- a chapter from a medical text (*Ayurveda*) corpus in Sanskrit
[2] Valmiki Ramayana -- full text of the epic *Ramayana* in English

Evaluation

Subjective Evaluation

- Survey (10 participants)
- Overall Rating 4.5

Objective Evaluation

- 26 Criteria
  - Technical
  - Functional
  - Data-related
- Performance
  - **Sangrahaka** (0.82)
  - FLAT (0.78)
  - WebAnno (0.74)
  - BRAT (0.70)

Technologies

https://github.com/hrishikeshrt/sangrahaka

Thank you!

Questions?