

Knowledge Management System Solution

KNOWLEDGE CHEAT TO KNOW

FABIEN FURFARO



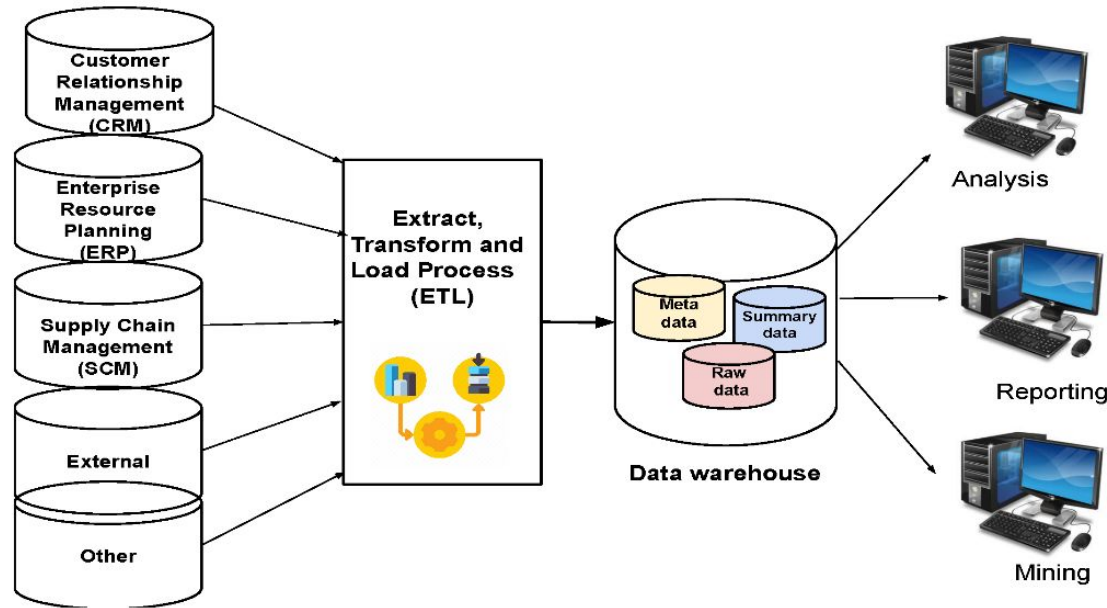
Définition : La connaissance est liée à la communication des données, c'est l'information stockée et structurée ?



Comment est gérée la connaissance aujourd'hui ?

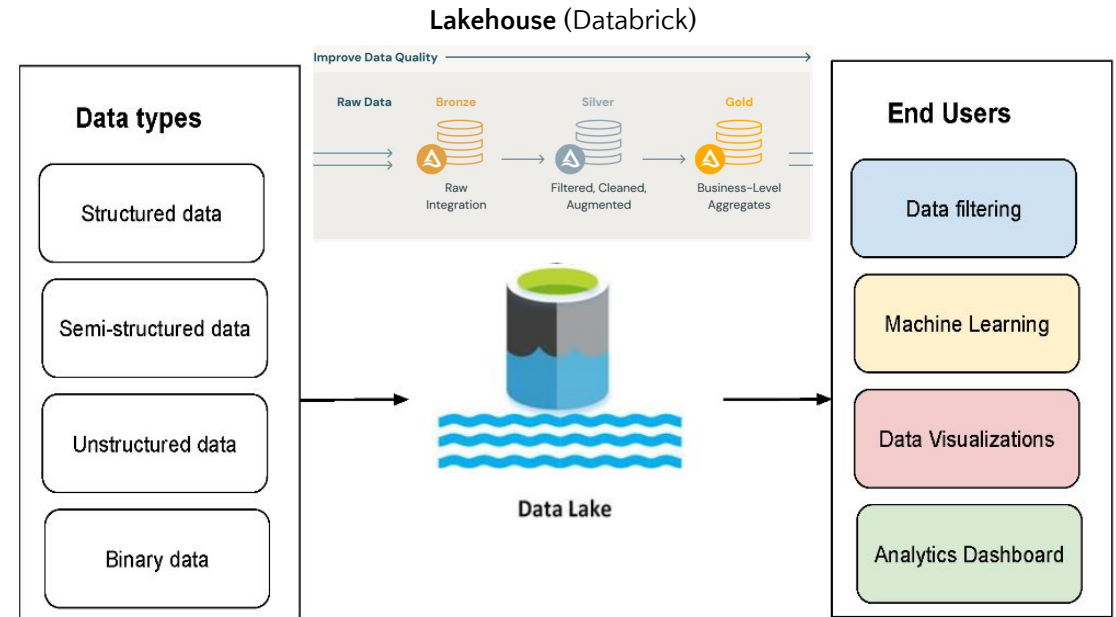
Pourquoi a-t-on besoin de gestion de la connaissance ?

Notes: La connaissance est lié à la communication



(a) Data warehouse

Invention de l'imprimerie
Accélération du partage de connaissance



(b) Data lake

Invention de l'imprimerie
Accélération du partage de connaissance

An Overview of Data Warehouse and Data Lake in Modern Enterprise Data Management (2022)

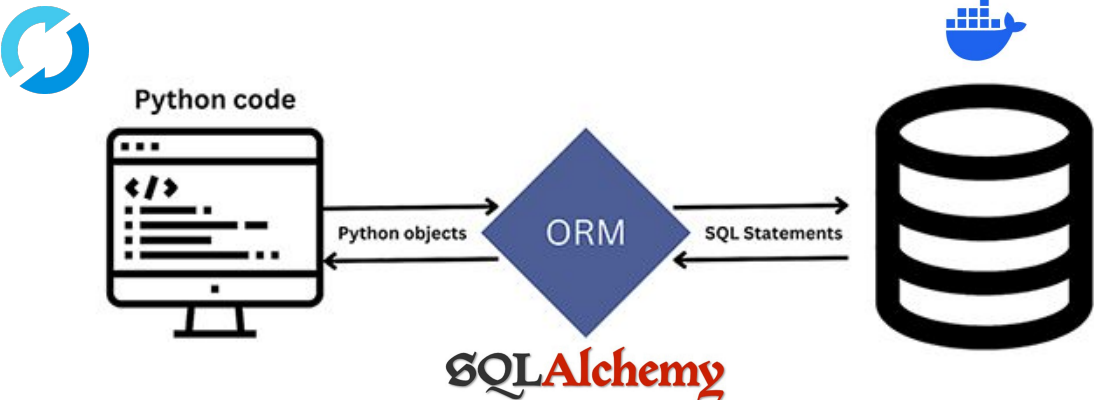
Comment optimiser la gestion des connaissances numériques ?



Le modèle de base de données relationnel

Pourquoi a-t-on besoin de gestion de la connaissance ?

Notes: La connaissance est lié à la communication (ACID)



Behaviour Driven Development
Acceleration du partage de connaissance

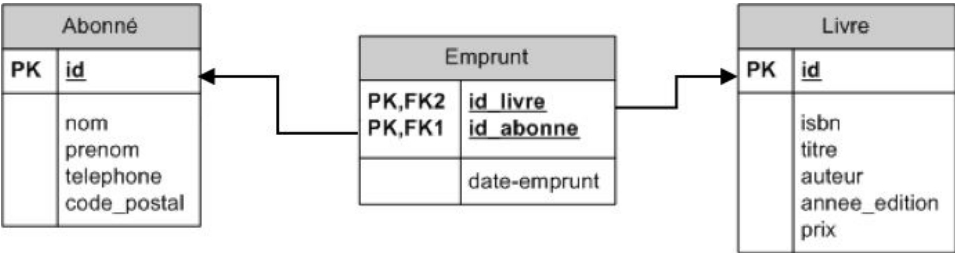
```
# app/models.py
from sqlalchemy import Column, Integer, String, ForeignKey
from sqlalchemy.ext.declarative import declarative_base
from sqlalchemy.orm import relationship

Base = declarative_base()

class Author(Base):
    __tablename__ = 'authors'
    id = Column(Integer, primary_key=True)
    name = Column(String)

class Book(Base):
    __tablename__ = 'books'
    id = Column(Integer, primary_key=True)
    title = Column(String)
    author_id = Column(Integer, ForeignKey('authors.id'))
    author = relationship('Author', back_populates='books')

Author.books = relationship('Book', order_by=Book.id, back_populates='author')
```



Affichez le nom et le prénom de tous les abonnés :

```
SELECT nom, prenom FROM ABONNE
```

Affichez tous les livres écrits par Voltaire :

```
SELECT * FROM LIVRE WHERE auteur='Voltaire'
```

Donnez le nom des abonnés ayant fait un emprunt :

```
SELECT A.nom
FROM ABONNE A INNER JOIN EMPRUNT E
ON A.id=E.id_abonne
```

Affichez les dates d'emprunt des exemplaires du livre "Le petit prince" :

```
SELECT E.date-emprunt
FROM EMPRUNT E INNER JOIN LIVRE L
ON E.id_livre=L.id
WHERE L.titre='Le petit prince'
```

Affichez le nombre de livres actuellement empruntés :

```
SELECT count(id_livre) FROM EMPRUNT
```

Affichez, pour chaque abonné ayant un emprunt en cours, le nombre de livres actuellement empruntés :

```
SELECT id_abonne, count(id_livre) FROM
EMPRUNT GROUP BY id_abonne
```

Affichez, pour chaque abonné ayant plus de 2 emprunts en cours, le nombre de livres actuellement empruntés :

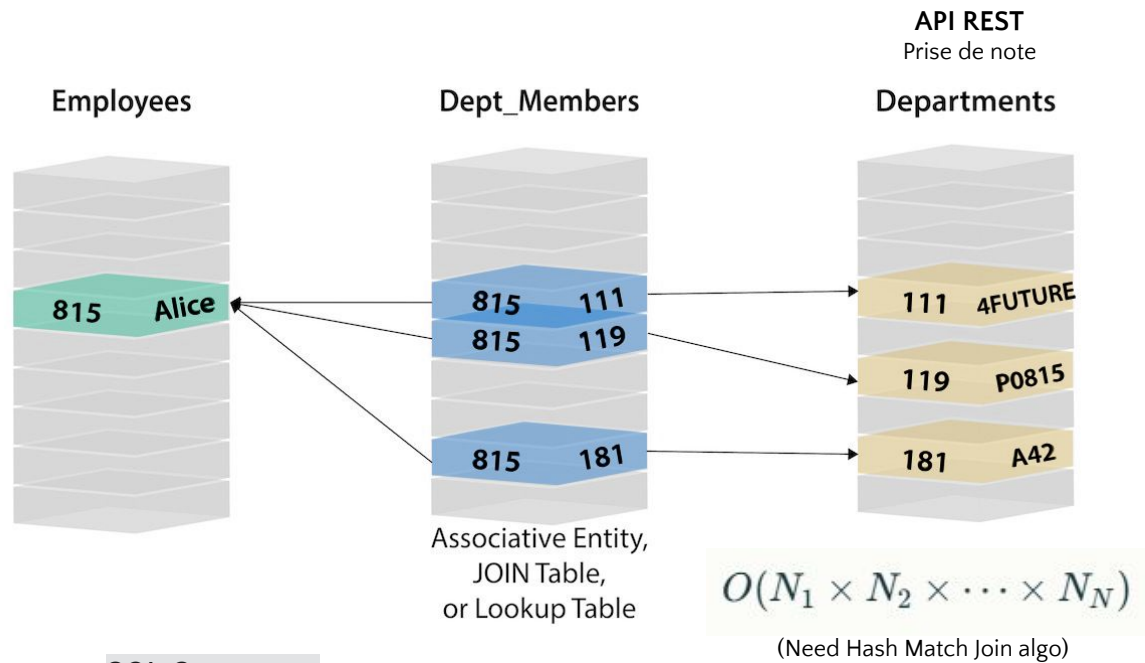
```
SELECT id_abonne, count(id_livre) FROM
EMPRUNT
GROUP BY id_abonne HAVING count(id_livre) >
2
```

Comment optimiser la gestion des connaissances numériques ?

Les limites du modèle relationnel

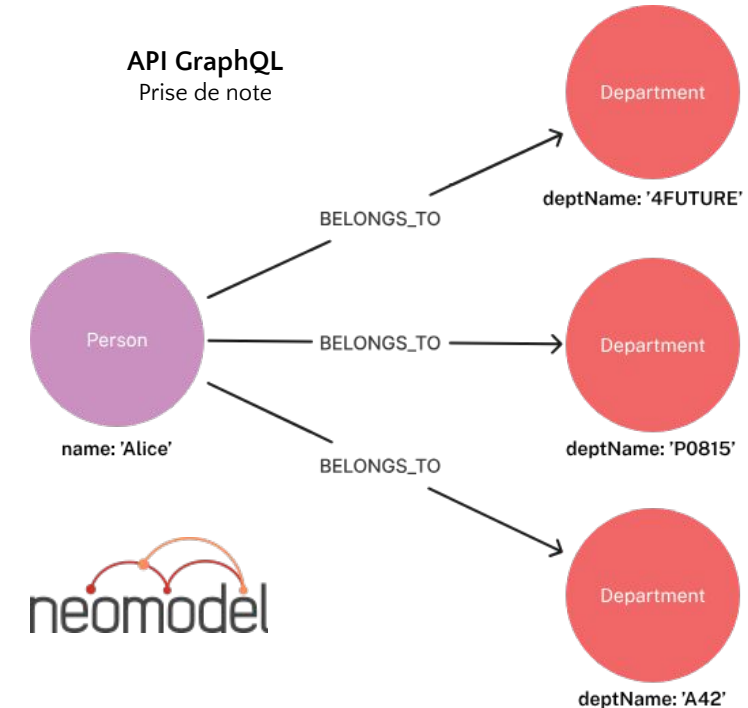
Pourquoi a-t-on besoin de gestion de la connaissance ?

Notes: La connaissance est lié à la communication (ACID)



SQL Statement

```
SELECT name FROM Person
LEFT JOIN Person_Department
  ON Person.Id = Person_Department.PersonId
LEFT JOIN Department
  ON Department.Id = Person_Department.DepartmentId
WHERE Department.name = "IT Department"
```



Cypher Statement

```
MATCH
(p:Person)-[:WORKS_AT]->(d:Dept)
WHERE d.name = "IT Department"
RETURN p.name
```

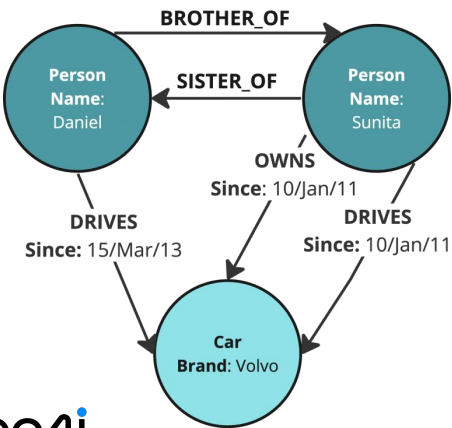
Comment optimiser la gestion des connaissances numériques ?



Le modèle de base données graphes

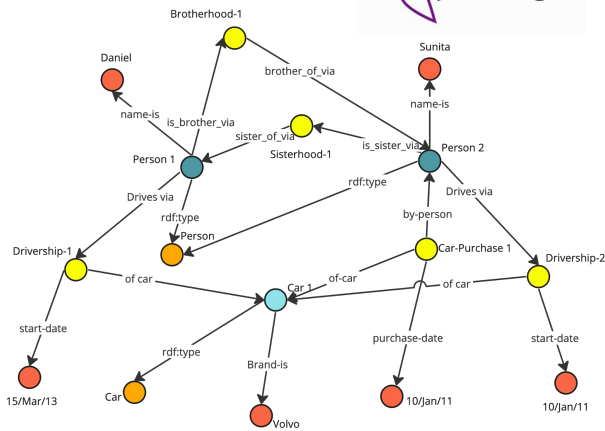
Pourquoi a-t-on besoin de gestion de la connaissance ?

Notes: La connaissance est lié à la communication



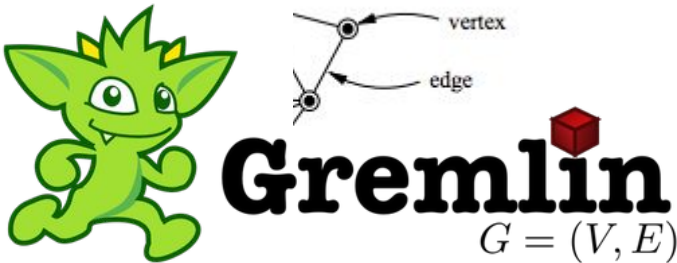
LPG

Label property (question driven)
(Neo4j/Cypher, Gremlin, etc.)



RDF

Web Sémantique (Ontologie modeling)
Subject predicate Object (OWL inference + parsing)
→ need SHACL engine (constraint validation)
+ editor&reasoner (Protégé axiom)



For each vertex in the graph, emit its label, then group and count each distinct label.

```
gremlin> g.V().label().groupCount()  
==>[occupation:21, movie:3883, category:18, user:6040]
```

What year was the oldest movie made?

```
gremlin> g.V().hasLabel('movie').values('year').min()  
==>1919
```

What is Die Hard's average rating?

```
gremlin> g.V().has('movie','name','Die Hard').inE('rated').values('stars').mean()  
==>4.121848739495798
```

Cypher	SPARQL
MATCH (js:Person)-[:KNOWS]-()-[:KNOWS]-(surfer) WHERE js.name = "Johan" AND surfer.hobby = "surfing" RETURN DISTINCT surfer	SELECT DISTINCT ?surfer WHERE { ?js a <Person>; <name> "Johan"; <knows>/<knows> ?surfer ; ?surfer <hobby> "surfing". }

Comment optimiser la gestion des connaissances numériques ?



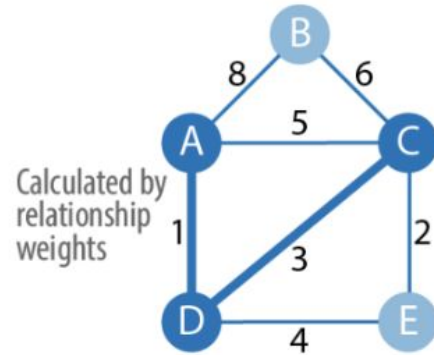
Les outils de la theorie des graphes

Pourquoi a-t-on besoin de gestion de la connaissance ?

Notes: La connaissance est lié à la communication

Graph Algorithm

Path finding, Community detection,
centrality...
→ use case



Calculated by
relationship
weights

Shortest Path

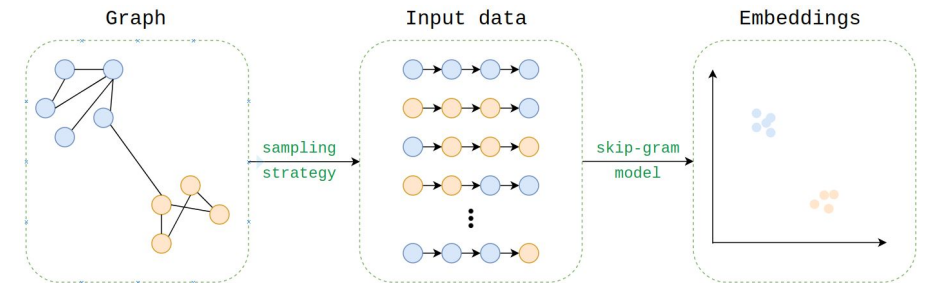
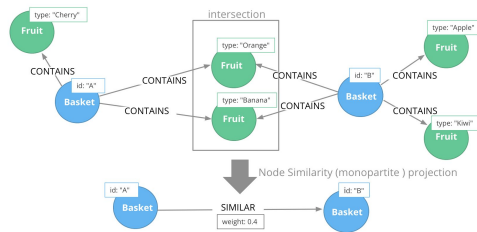
Shortest path between
2 nodes (A to C shown)

Cypher Statement

```
MATCH
(p:Person)-[:WORKS_AT]->(d:Dept)
WHERE d.name = "IT Department"
RETURN p.name
```

Graph DataScience

Node classification, similarity, Node2vec



Cypher Statement

```
MATCH
(p:Person)-[:WORKS_AT]->(d:Dept)
WHERE d.name = "IT Department"
RETURN p.name
```

Cypher Statement

```
MATCH
(p:Person)-[:WORKS_AT]->(d:Dept)
WHERE d.name = "IT Department"
RETURN p.name
```

Comment optimiser la gestion des connaissances numériques ?



D'autres outils que les graphes ? (construire connaissance, ex: NLP)

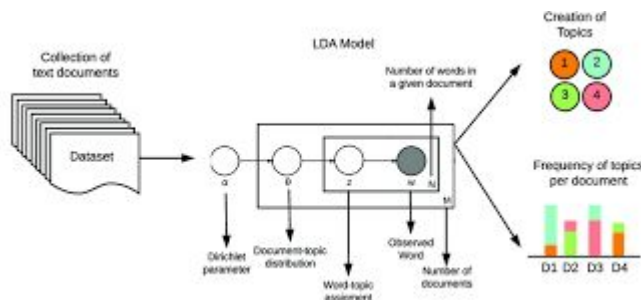
Pourquoi a-t-on besoin de gestion de la connaissance ?

Notes: La connaissance est liée à la communication

$$w_{x,y} = tf_{x,y} \times \log\left(\frac{N}{df_x}\right)$$

Tools (entity extraction and Topic modeling)

BoW (TF) → LDA (pyLDAvis)
Spacy (+textacy), NLTK, Gensim, YoLo
→ Data Mining



Cypher Statement

```
MATCH
(p:Person) -[:WORKS_AT]->(d:Dept)
WHERE d.name = "IT Department"
RETURN p.name
```

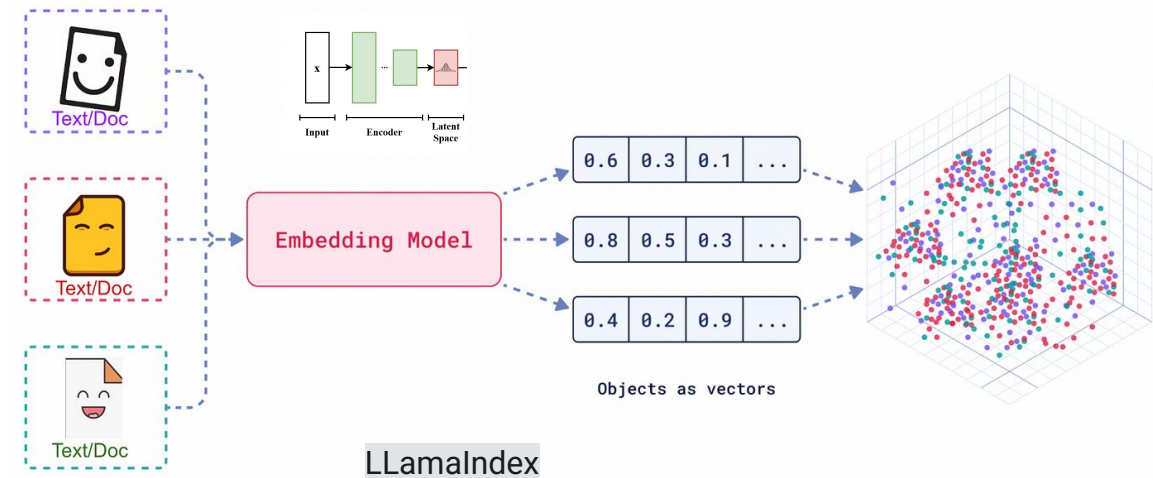
Cypher Statement

```
MATCH
(p:Person) -[:WORKS_AT]->(d:Dept)
WHERE d.name = "IT Department"
RETURN p.name
```

In fact, the **Chinese** market has the **three** most influential names of the retail and tech space – **Alibaba**, **Baidu**, and **Tencent** (collectively touted as **BAT**), and is betting big in the global **AI** in retail industry space. The **three** giants which are claimed to have a cut-throat competition with the **U.S.** (in terms of resources and capital) are positioning themselves to become the future **AI** platforms. The trio is also expanding in other **Asian** countries and investing heavily in the **U.S.** based **AI** startups to leverage the power of **AI**. Backed by such powerful initiatives and presence of these conglomerates, the market in APAC AI is forecast to be the fastest-growing **one**, with an anticipated **CAGR** of **45%** over **2018 - 2024**.

To further elaborate on the geographical trends, **North America** has procured **more than 50%** of the global share in **2017** and has been leading the regional landscape of **AI** in the retail market. The **U.S.** has a significant credit in the regional trends with **over 65%** of investments (including M&As, private equity, and venture capital) in artificial intelligence technology. Additionally, the region is a huge hub for startups in tandem with the presence of tech titans, such as **Google**, **IBM**, and **Microsoft**.

$$S_C(A, B) := \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|}$$



LLamaIndex

```
MATCH
(p:Person) -[:WORKS_AT]->(d:Dept)
WHERE d.name = "IT Department"
RETURN p.name
```

Comment optimiser la gestion des connaissances numériques ?

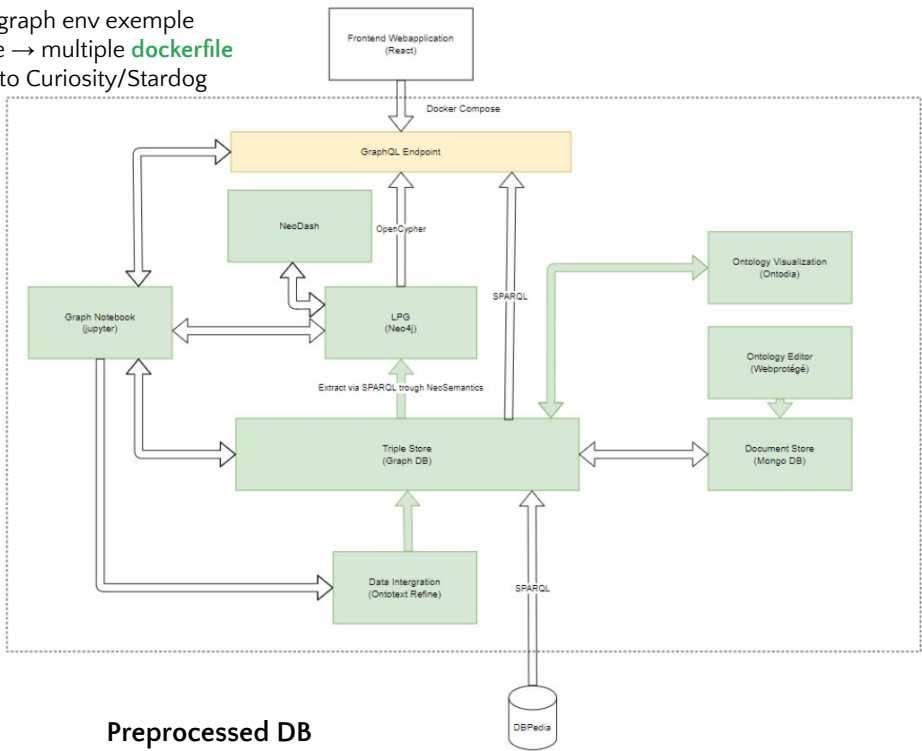
Les architectures de gestion de connaissance

Pourquoi a-t-on besoin de gestion de la connaissance ?

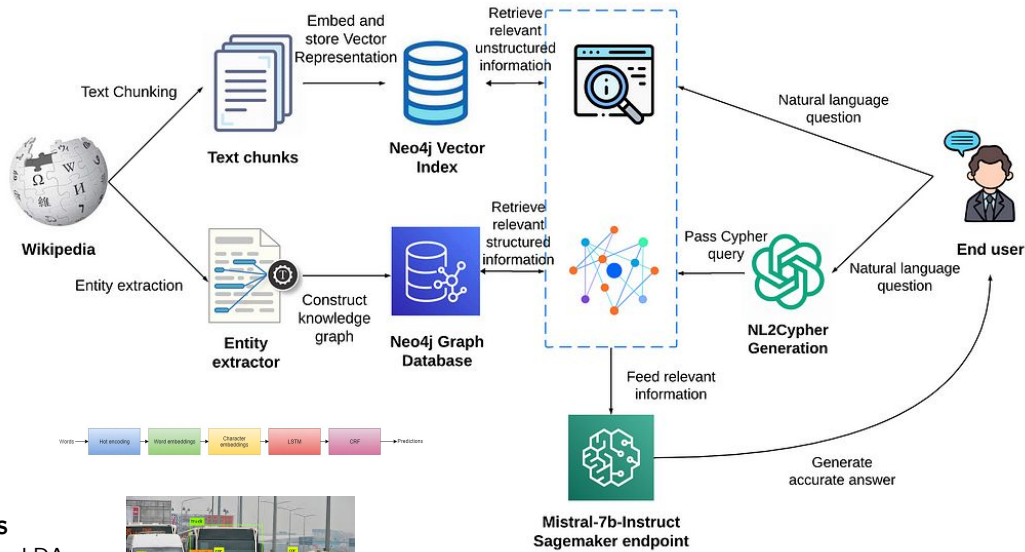
Notes: La connaissance est lié à la communication

Knowledge Graph = Data + Relations + Semantic Structure

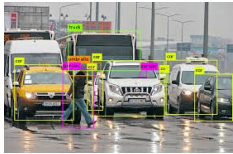
Workbench Architecture
Locla hybrid graph env exemple
Docker-compose → multiple **dockerfile**
It's equivalent to Curiosity/Stardog



Preprocessed DB
Need EDA good practice to construct KG



Tools
BoW (TF) → LDA
Spacy, NLTK,
Gensim, YoLo



Business Case
exemple

- Problem statement
- Mission Statement
- Business Model Canva
- Intended Impact

Construction of Knowledge Graphs: State and Challenges (2023)

Comment optimiser la gestion des connaissances numériques ?



About Capgemini

Capgemini is a global leader in partnering with companies to transform and manage their business by harnessing the power of technology. The Group is guided everyday by its purpose of unleashing human energy through technology for an inclusive and sustainable future. It is a responsible and diverse organization of 270,000 team members in nearly 50 countries. With its strong 50 year heritage and deep industry expertise, Capgemini is trusted by its clients to address the entire breadth of their business needs, from strategy and design to operations, fuelled by the fast evolving and innovative world of cloud, data, AI, connectivity, software, digital engineering and platforms. The Group reported in 2020 global revenues of €16 billion.

Get the Future You Want | www.capgemini.com



This presentation contains information that may be privileged or confidential and is the property of the Capgemini Group.

Copyright © 2021 Capgemini. All rights reserved.