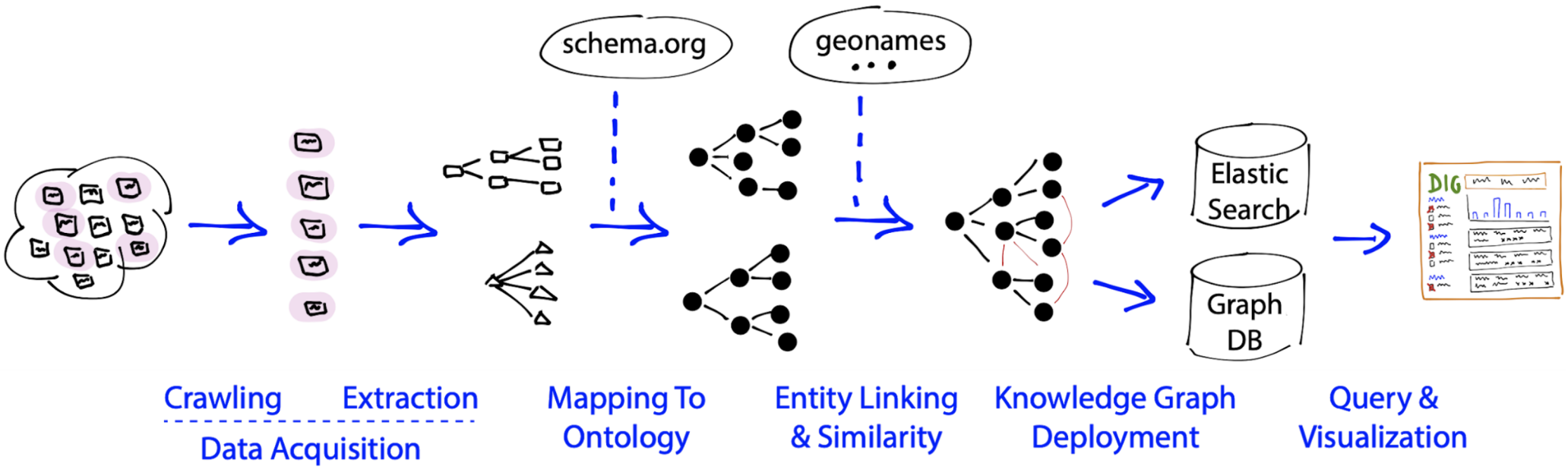
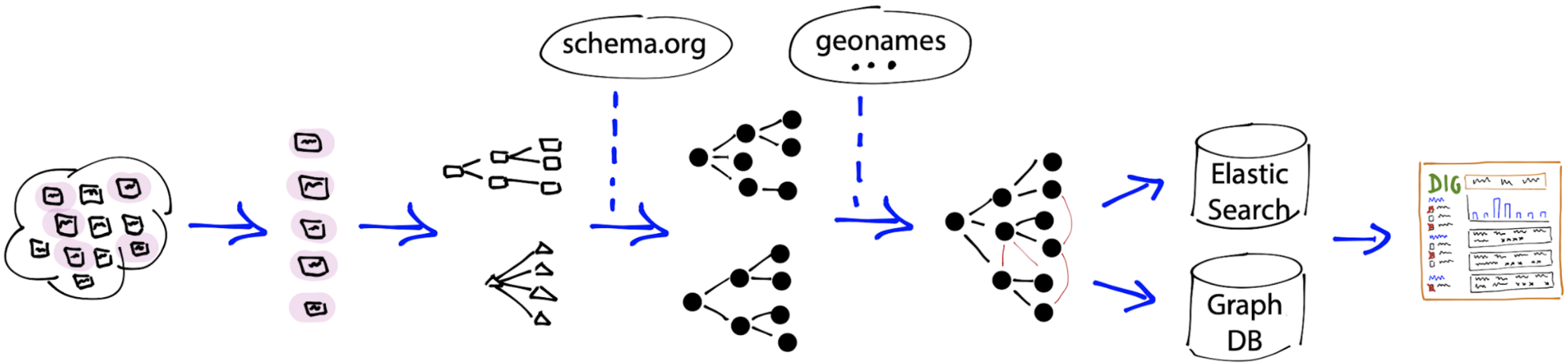
USC Viterbi
School of Engineering

Course Overview

Craig Knoblock, Jay Pujara & Minh Pham

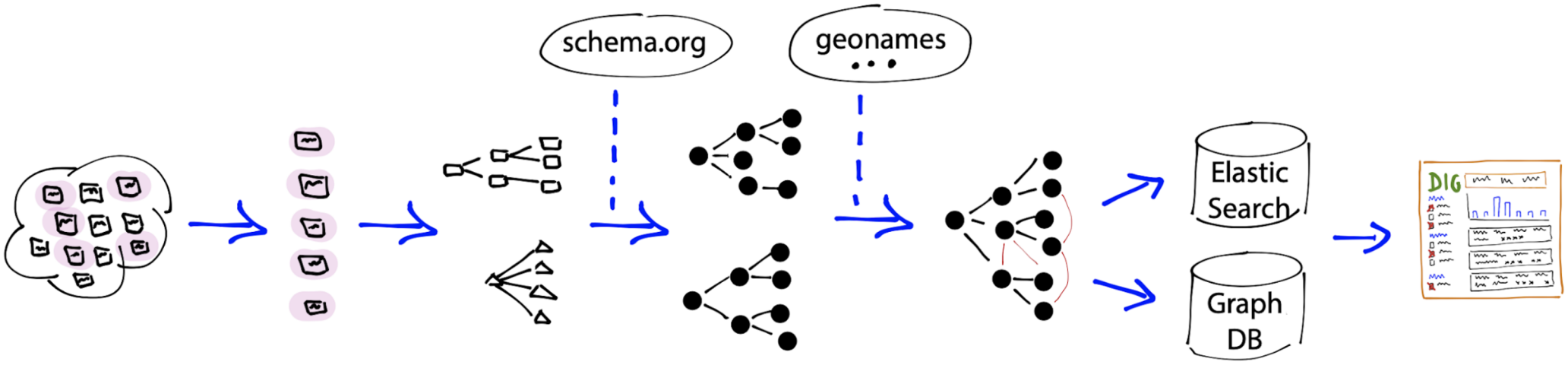

schema.org
geonames
Elastic Search
Graph DB
DIG
Crawling
Extraction
Data Acquisition
Mapping To Ontology
Entity Linking & Similarity
Knowledge Graph Deployment
Query & Visualization



Crawling &

Intellectual property

Information extraction



Crawling Extraction

Data Acquisition

Mapping To Ontology

Entity Linking & Similarity

Knowledge Graph Deployment

Query & Visualization

Crawling &

Intellectual property

Information extraction

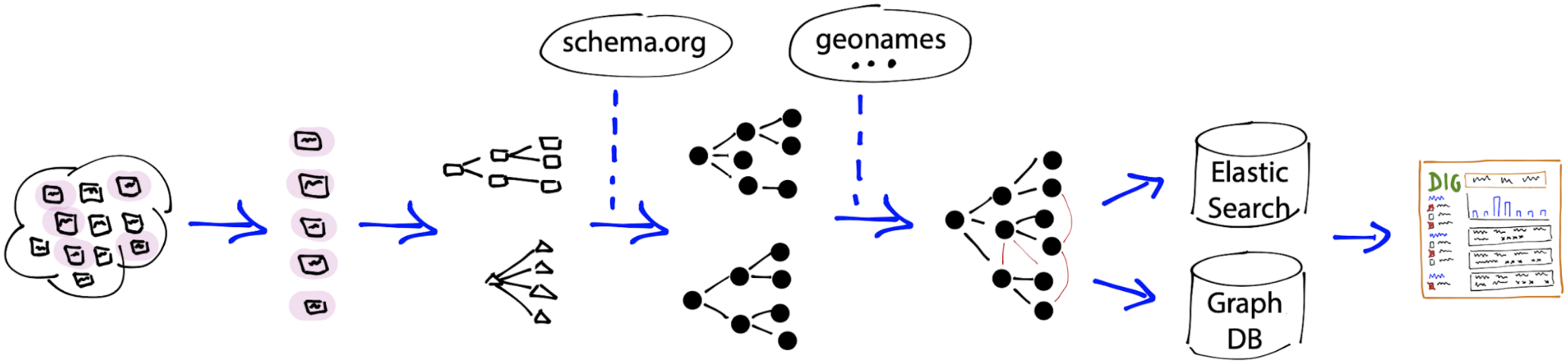
RDF

Ontologies

Semantic Labeling & Modeling

Large KGs

Linked Data & Semantic Web



Crawling Extraction

Data Acquisition

Mapping To Ontology

Entity Linking & Similarity

Knowledge Graph Deployment

Query & Visualization

Crawling &

Intellectual property

Information extraction

RDF

Ontologies

Semantic Labeling & Modeling

Large KGs

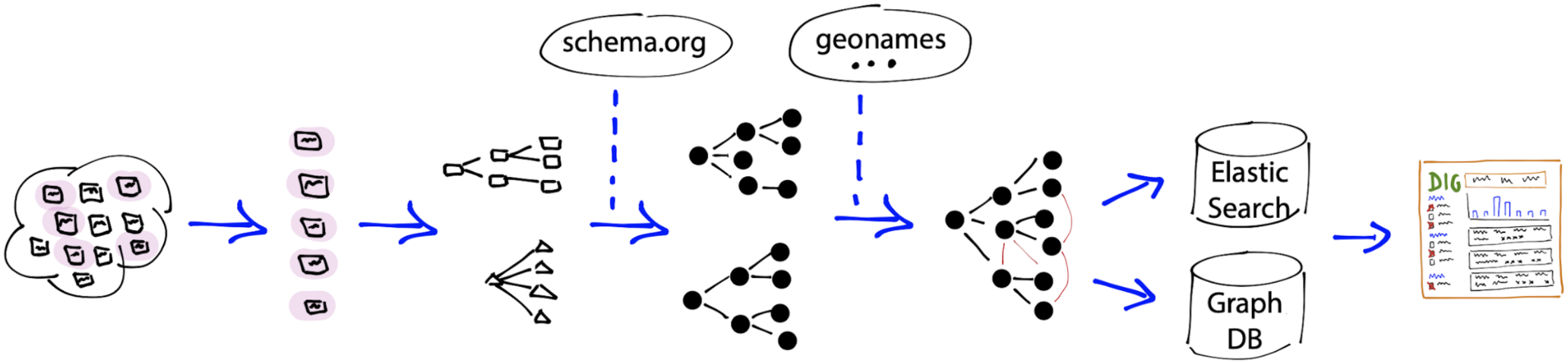
Linked Data & Semantic Web

String similarity

ER & PSL

Blocking & Relational ER

KG embeddings



Crawling Extraction

Data Acquisition

Mapping To Ontology

Entity Linking & Similarity

Knowledge Graph Deployment

Query & Visualization

Crawling &

Intellectual property

Information extraction

RDF

Ontologies

Semantic Labeling & Modeling

Large KGs

Linked Data & Semantic Web

String similarity

ER & PSL

Blocking & Relational ER

KG embeddings

Queries and KGs

KG embeddings

KG use cases

Graph Analytics

Question Answering

Linked Maps

Common Sense Knowledge

Slide 1

- Intro [JP]
  - What is a KG?
  - How are KGs used?
  - Where do KGs come from?
- QS: Crawling the Web [CK]
  - Surface web vs. deep web vs. dark web
  - What are the challenges?
  - What is the basic architecture of a crawler?

Slide 2

- QS: Information Extraction [JP]
  - What are the three components to IE?
  - How do they differ with differing amounts of supervision?
  - How do real-world systems do IE?
- QS: Knowledge Representation [CK]
  - What are the basic elements of RDF?
  - What are the different syntaxes for RDF?
  - What is RDF Schema?

Slide 3

- QS: Entity Resolution [JP]
  - What are the three basic settings of ER?
  - What are real-world examples of these ER settings?
  - What predictive techniques are used for ER?
- QS: Large KGs [CK]
  - Which large KGs exist and what distinguishes them?
  - How is Wikidata different from other large KGs?
  - What are the different methods to get data from large KGs?

Slide 4

- QS: Queries & KGs [JP]
  - SPARQL syntax
  - Graph query structure
- Large KGs – Entity Linking [CK]
  - What is entity linking and why is it hard?
  - What is the basic architecture of entity linkers?
  - What are the popular methods for disambiguation?

Slide 5

- String Similarity [CK]
  - What is the difference between token and sequence based?
  - What are the strengths and weaknesses of each method?
  - How do hybrid methods work?
- Information Extraction [JP]
  - What are labeling functions (Snorkel)? What makes them good?
  - What does the generative model do?
  - What is the purpose of the discriminative model?

Slide 6

- ER & Probabilistic Soft Logic (PSL) [JP]
  - PSL fundamentals and when collective inference is necessary
  - Logic syntax and computing a soft logic loss
  - PSL model for knowledge graph identification
- Ontologies & RDF [CK]
  - What are ontologies for?
  - What are the building blocks of OWL?
  - How to represent knowledge in OWL?

Slide 7

- Structured Data [JP]
  - Different patterns and layouts of structured data
  - How to extract knowledge from structured datasets
  - DataCube knowledge representation
- Semantic Labeling/Semantic Models [CK]
  - What is semantic labeling (SL)?
  - How do rule-based and learning based SL approaches work?
  - What are the steps to discover and build the semantic models of data sources ?

Slide 8

- Knowledge Graph Embeddings [JP]
  - What are KG embeddings for?
  - What are the popular methods for KG embeddings?
  - What are the advantages, disadvantages of different methods?
- Data Cleaning [MP]
  - What are the three main challenges in data cleaning?
  - What are the popular methods for error detection ?
  - What are the popular methods for error repairing/data transformation ?

Slide 9

- Blocking and Relational ER [JP]
  - Why we need blocking
  - Techniques used for blocking (LSH)
  - Relational entity resolution patterns / ER constraints
- Linked Data & Semantic Web [CK]
  - What are the principles of linked data?
  - How is DBpedia constructed?
  - What were the challenges in building the linked for the AAC?

Slide 10

- Graph Analytics [JP]
  - Node importance / centrality
  - Clustering measures and techniques
- Special Topics: Linked Maps & Semantic Modeling [CK]
  - How to represent geospatial data as linked data ?
  - How to use known semantic models to model a new data source ?
  - How to use knowledge graphs to model a new data source ?

Slide 11

- Special Topics: Commonsense Knowledge Graphs [JP]
  - How is commonsense knowledge different from other knowledge?
  - What are the main KGs for commonsense knowledge?
- Question Answering [JP]
  - QA challenges
  - Watson case study - focus entities, lexical types, question types
  - Semantic labeling and SPARQL query discovery

Slide 12

- Intellectual Property [CK]
  - What mechanisms exist to protect intellectual property?
  - What are the requirements for a patent?
  - What can be copyrighted, and what is the difference in CC licenses?