



香港科技大學
THE HONG KONG UNIVERSITY OF
SCIENCE AND TECHNOLOGY



THE DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Knowledge Base Refinement and Enhancement

Lei Chen


The Hong Kong University of Science and Technology


Outline

- **Background**
- **Overview of Current Challenges**
 - Knowledge Base Construction
 - Knowledge Base Refinement
 - Completion (incompleteness)
 - Error detection (incorrectness)
- **Our Works**
- **Conclusions**

Background: What is a knowledge base?

- A knowledge base stores a collection of *facts* in the form of <**subject**, relation, object> triple

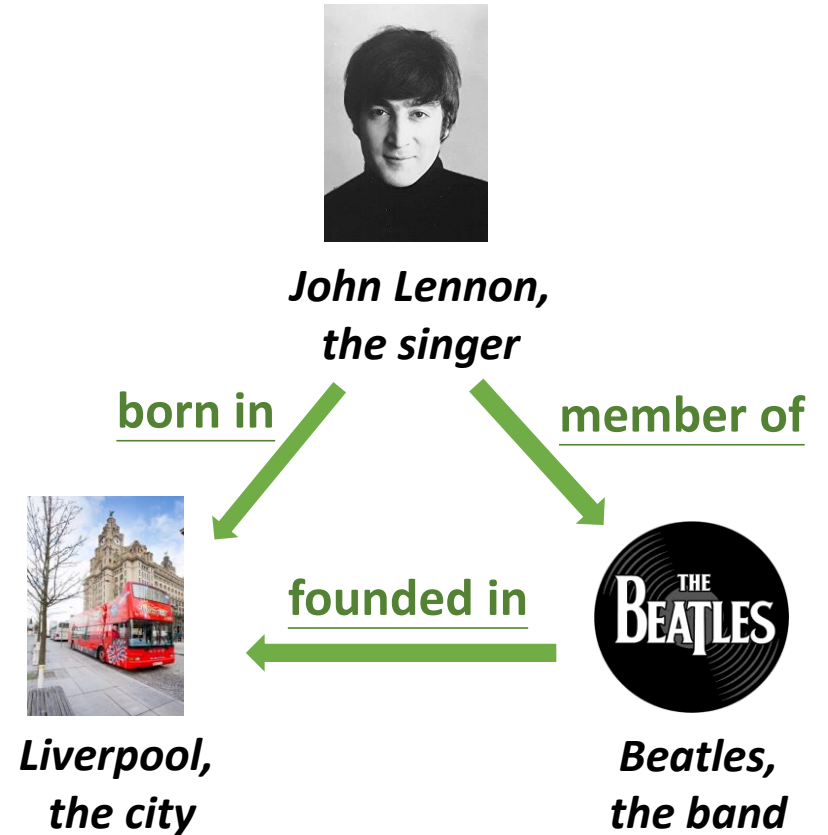

real-world entity


real-world entity or literal

< **John Lennon**, member Of, Beatles >

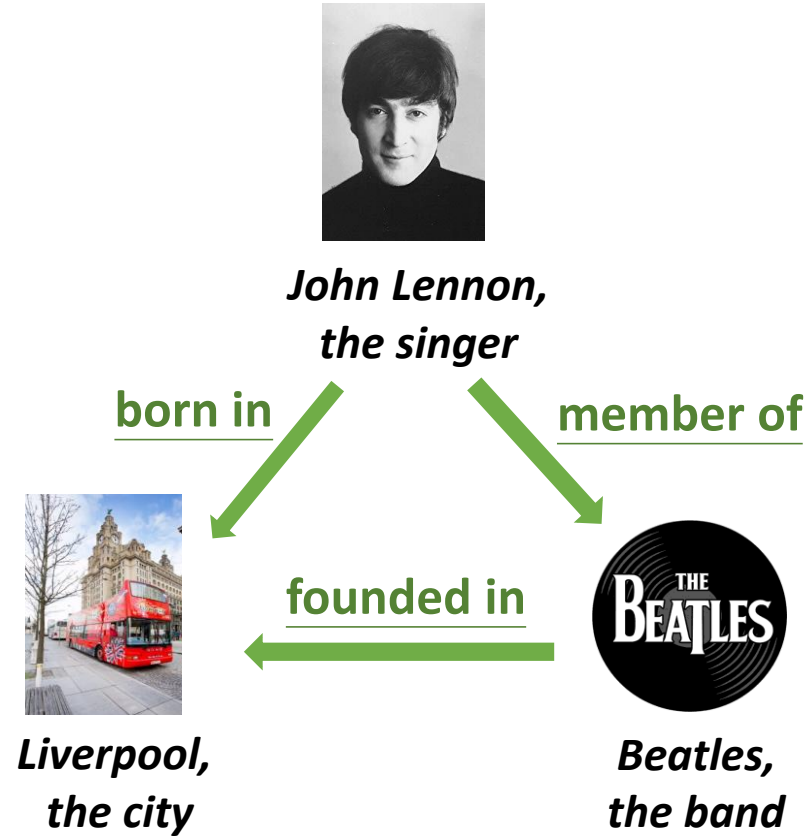
< **John Lennon**, born In, Liverpool >

< **Beatles**, founded In, Liverpool >



Background: Why knowledge bases?

- Entity searching, Question Answering, Recommendation...



where was John Lennon born

All Images Maps News Videos More Settings Tools

About 24,900,000 results (0.79 seconds)

John Lennon / Place of birth

Liverpool, United Kingdom

Background: Where do knowledge bases come from?

- Structured Text
 - *Cyc, Freebase, Wikidata*
- Semi-structured Data
 - *DBpedia, Yago*
- Unstructured Text
 - *Nell, Knowledge Vault, Probase*

Beatles last live performance

Published: Thursday, January 26th 2017, 5:24 am PST

Updated: Monday, January 30th 2017, 4:06 am PST

Written by Jim Eftink, Producer [CONNECT](#)



(KFVS) - How about a little Beatles history.

It was on this date in 1969, the band performed their last live public performance.

Allan Williams, First Manager of the Beatles, Dies at 86

(Source: Stock images) By ALLAN KOZINN DEC. 31, 2016

The Beatles
January 17 at 10:00am · 🌐

The Harrison family is proud to announce the release of George Harrison – The Vinyl Collection box set featuring all of George Harrison's solo studio albums in one collection for the first time.

GEORGE HARRISON – THE VINYL COLLECTION
Released on 24th February, 2017, the vinyl box set includes all twelve of George's studio albums with exact replicas of the original release track listing and artwork. Also included in the box set are George's classic live album Live in Japan (2x) ... [See More](#)

George Harrison - The Vinyl Collection - Released February 24th 2017
George Harrison - The Vinyl Collection, available to pre-order now with an exclusive & limited edition... [YOUTUBE.COM](#)

Like Love would 3K Comment Share

900 shares

Write a comment...

Jeffrey Smith What I would really be interested in is an "All Things Must Pass...Stripped Down" with just the basic tracks without Phil Spector's Wall of Sound. It has to be said it would sound really good and I would buy it in a heartbeat.
Like · Reply · 88g31 January 17 at 10:25am

17 Replies

Dave Standing I can just see the greedy Harrison family and the greedy music industry multi-millionaire big wigs rubbing their hands with glee once more whilst discussing various methods to make people buy their already bought and paid for record collections all at once... [See More](#)
Like · Reply · 88g29 January 17 at 10:15am · Edited

30 Replies

View more comments 2 of 100

The Beatles
January 17 at 8:58am · 🌐

"Of very few individual songs can it be said, 'This changed the course of popular music.' 'A Day in The Life' is one such song." – Richard Havers

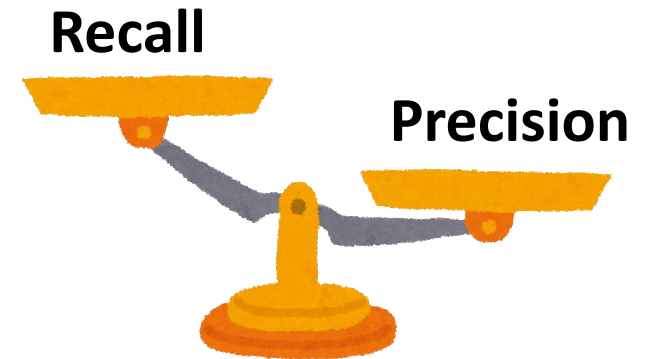
The Beatles - A Day In The Life
A Day in The Life The Beatles 1 Video Collection is Out Now. Get your copy here.



anager of the Beatles in 1960, he sent them on a stint in Germany tagecraft. Press Association, via Associated Press

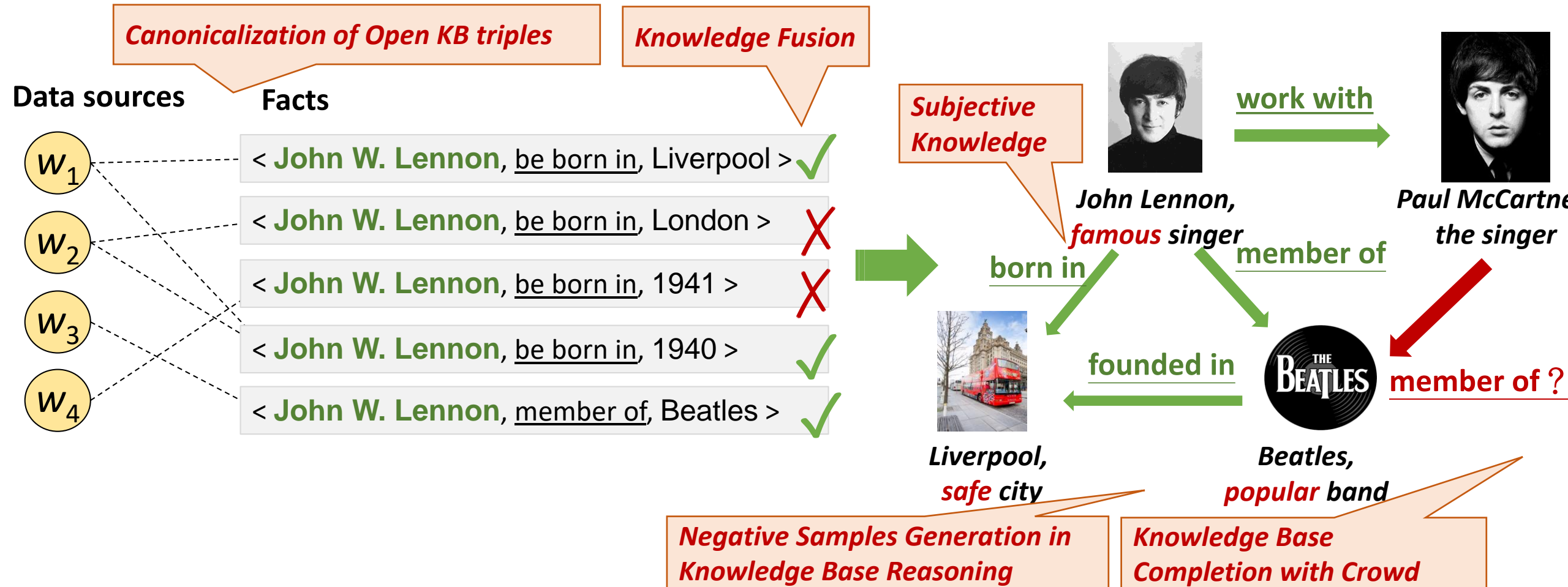
Current Problems in Knowledge Bases

- **Incompleteness**
 - Most KBs can hardly cover all the entities in the real world
- **Incorrectness**
 - There may be errors during KB construction
- **Domain-specific information**
 - It is hard to build a complete and sound KB in certain domains
 - It is hard to involve subjective information in KBs



Summary:

Our Recent Works about Knowledge Bases



Our Recent Works about Knowledge Bases

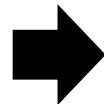
Work: Canonicalization of Open Knowledge Bases

- Our proposed method: [Lin, ICDE 2019]
 - Model the canonicalization problem of noun phrases and relation phrases in the Open KB triples *jointly*
 - Utilize the *side information* obtained from the source text of the triples

< **Peter Cook**, hosted, “Not Only... But Also” >



Peter Cook hosted The TV series “**Not Only ... But Also**”. **John Lennon**, born in **Liverpool**, one of the members in **Beatles**, had a pleasant cooperation with **Cook** in this show.



• *Candidate Entities in the Source Text*



...

- *Type Lists of These Entities*
- *The Domain Vector of the Source Text*

Our Recent Works about Knowledge Bases

Work: Subjective Knowledge Base Construction

- Our proposed method: **[Xin, SIGMOD 2018]**

Input: Existing KBs + Crowd

Goal: Enrich KB with subjective information



*Liverpool,
a city*



*John Lennon,
a singer*



*Beatles,
a band*



*Liverpool,
safe city*



*John Lennon,
famous singer*



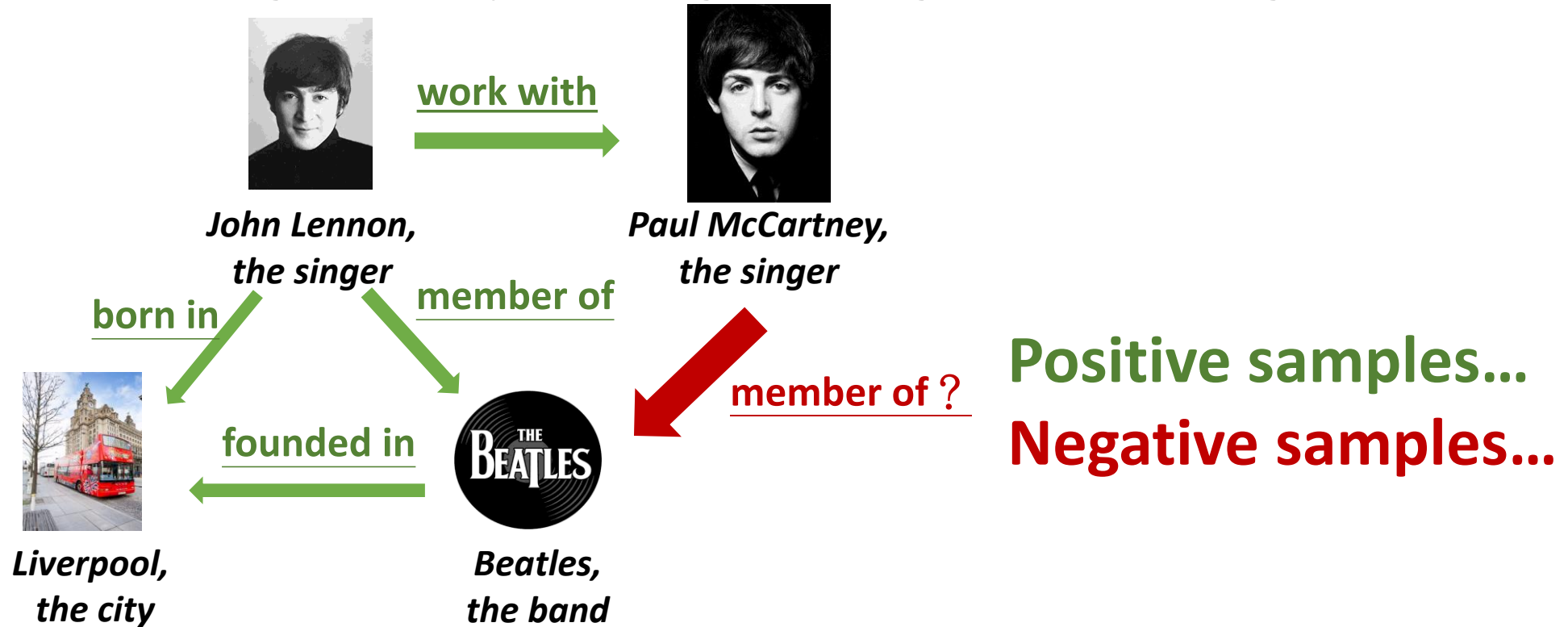
*Beatles,
popular band*

[Xin, SIGMOD 2018] Xin, Hao, Rui Meng, and Lei Chen. "Subjective Knowledge Base Construction Powered By Crowdsourcing and Knowledge Base." *Proceedings of the 2018 International Conference on Management of Data (SIGMOD)*. ACM, 2018.

Our Recent Works about Knowledge Bases

Work: Knowledge Base Reasoning [Zhang, ICDE 2019]

Goal: Generate negative samples during knowledge base reasoning

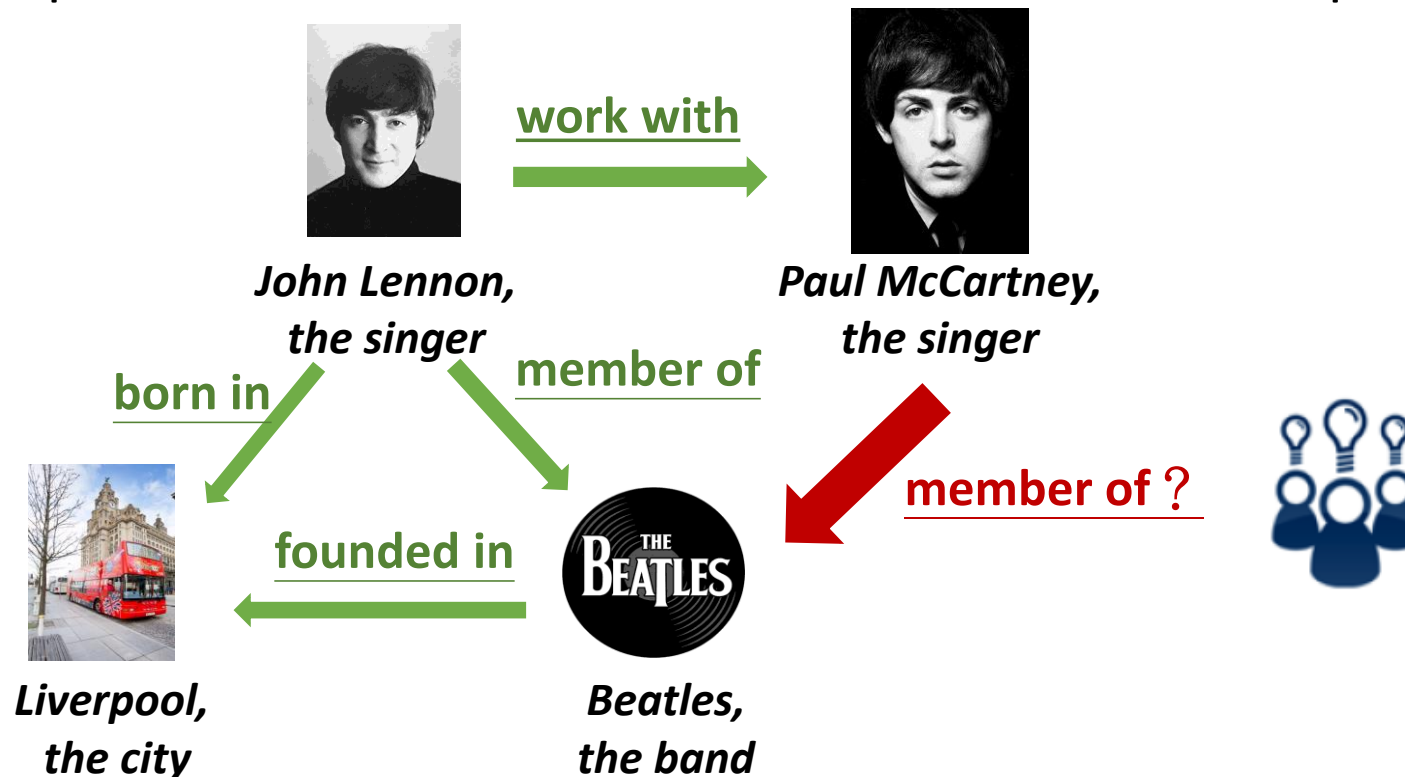


[Zhang, ICDE 2019] Zhang, Yongqi, Quanming Yao, Yingxia Shao, and Lei Chen. " NSCaching: Simple and Efficient Negative Sampling for Knowledge Graph Embedding." 2019 *IEEE 35th International Conference on Data Engineering (ICDE)*. IEEE, 2019.

Our Recent Works about Knowledge Bases

Work: Knowledge Base Enhancement [Jiang, ICDE 2018]

Goal: Update the current KBs with new facts with the help of the crowd



[Jiang, ICDE 2018] Jiang, Linnan, Lei Chen, and Zhao Chen. "Knowledge Base Enhancement via Data Facts and Crowdsourcing." *2018 IEEE 34th International Conference on Data Engineering (ICDE)*. IEEE, 2018.

Our Recent Works about Knowledge Bases

Work: Knowledge Fusion [Lin. VLD 2018]

Input: Different **sources** providing conflicting **facts**

Goal: Discover trustworthy information (i.e., the **truths**)



[Lin, VLDB 2018] Lin, Xueling, and Lei Chen. "Domain-aware multi-truth discovery from conflicting sources." *Proceedings of the VLDB Endowment* 11.5 (2018): 635-647.

Conclusions

- Knowledge base construction needs to be improved
 - **Entity Extraction and Entity Resolution**
 - **Relation Extraction**
- Knowledge bases need refinement and enhancement (internal methods vs external methods)
 - **Incompleteness**
 - Hardly can cover all the entities/relations in the real world
 - How to add new knowledge, e.g., subjective information, to the current KB?
 - **Incorrectness**
 - May have errors when constructing the KB

Future Works

- **Possible research directions:**
 - Knowledge Graph Embedding (for Knowledge Reasoning in KB Completion and Error Detection)
 - Knowledge Fusion
 - Open Information Extraction Refinement
 - Knowledge Base Refinement via Human Power

References

- **[Lin, VLDB 2020]** "KBPearl: a Knowledge Base Population System Supported by Joint Entity and Relation Linking.", Proceedings of the **VLDB** Endowment 13, no. 7 (2020): 1035-1049.
- **[Zhang, ICDE 2020]**, Quanming Yao, Wenyuan Dai, Lei Chen. "AutoSF: Searching Scoring Functions for Knowledge Graph Embedding.", In 2020 IEEE 36th International Conference on Data Engineering (**ICDE**). IEEE, 2020
- **[Zhang, ICDE 2019]** Zhang, Yongqi, Quanming Yao, Yingxia Shao, and **Lei Chen**. "NSCaching: Simple and Efficient Negative Sampling for Knowledge Graph Embedding." 2019 *IEEE 35th International Conference on Data Engineering (ICDE)*. IEEE, 2019.
- **[Jiang, ICDE 2018]** Jiang, Linnan, **Lei Chen**, and Zhao Chen. "Knowledge Base Enhancement via Data Facts and Crowdsourcing." 2018 *IEEE 34th International Conference on Data Engineering (ICDE)*. IEEE, 2018.
- **[Lin, VLDB 2018]** Lin, Xueling, and **Lei Chen**. "Domain-aware multi-truth discovery from conflicting sources." *Proceedings of the VLDB Endowment* 11.5 (2018): 635-647.
- **[Lin, ICDE 2019]** Lin, Xueling, and **Lei Chen**. "Canonicalization of Open Knowledge Bases with Side Information from the Source Text." 2019 *IEEE 35th International Conference on Data Engineering (ICDE)*. IEEE, 2019.
- **[Xin, SIGMOD 2018]** Xin, Hao, Rui Meng, and **Lei Chen**. "Subjective Knowledge Base Construction Powered By Crowdsourcing and Knowledge Base." *Proceedings of the 2018 International Conference on Management of Data (SIGMOD)*. ACM, 2018.