Combining NER and Knowledge Graphs in the Heritage Connector Project

Kalyan Dutia

SCIENCE MUSEUM GROUP

UKRI Arts and Humanities Research Council

SCHOOL OF ADVANCED STUDY UNIVERSITY OF LONDON

V&A

How can existing digital tools and methods be used to build relationships at scale between poorly and inconsistently catalogued digitised collection objects and other content sources?

Is such an approach scalable to larger volumes of content and different types of collections?

Where is the best use of human input in supporting such an approach? What expertise and skills are required for this input?

ai4lam | 2021-02-16 ai4lam Community | Sokol space suit | Science Muse

https://collection.sciencemuseumgroup.org.uk/objects/co8538105/sokol-space-suit-space-suit

Search objects, people, categories, object numbers etc.

Sokol space suit

MADE: 1991 in Russia

British astronaut, Helen Sharman's Sokol spacesuit made by Zvezda. Sharman wore this rescue suit during the space flight on board the SOYUZ-TM-12 and MIR spacecraft in May 1991. Space suit model number KV-2 No. 167.

Sokol-KV-2 rescue suit worn by Helen Sharman during the Juno mission to the Mir space station, 1991

СПАСАТЕЛЬНЫЙ СКАФАНДР

Helen Sharman was the first British person in space. Sharman wore this suit for two hours on the ground to check its fit. Lying back, she tried to read but her arms ached from holding the book for so long. Despite the suit's cooling systems she sweated 2 litres during the mission launch. Once she could remove the suit, she dried it thoroughly to ensure it would not go mouldy.

The Sokol suit was developed after three unsuited cosmonauts asphyxiated on the Soyuz 11 mission in 1971 when their descent module depressurised during the return to Earth. Every cosmonaut now wears one during launch and return from space. It will keep the wearer alive for a number of hours in the event of a cabin depressurisation. Each suit is tailor made to the individual cosmonaut and comprises an inner, airtight 'bladder' of rubberised plastic and an outer layer of nylon canvas. There are connecting rings on the lower abdomen for air (cooling) and oxygen supplies and a centrally positioned pressure adjustment valve control on the chest; the pressure gauge is on the left wrist. The helmet and boots are integral with the rest of the suit; the gloves are attached with anodized aluminium bayonet fixings. Today's Sokol design is little changed from the original.

Source: Zvezda

ON DISPLAY

Science Museum: Exploring Space Gallery

If you are visiting to see this object, please contact us in advance to make sure that it will be on display.

RELATED PEOPLE

Helen Sharman

RELATED ARTICLES

National Science and Media Museum

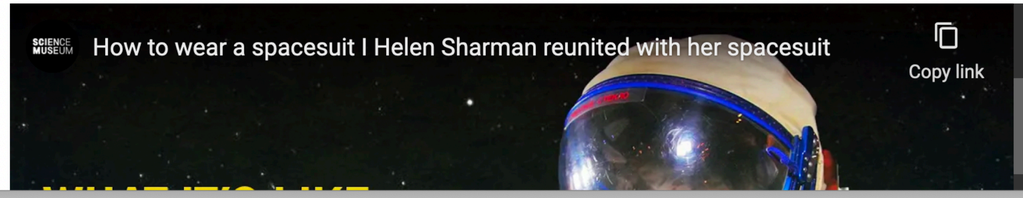
Bring the National Science and Media Museum collection home in Animal Crossing

Science Museum

Highlights on display

Science Museum announces National Lottery ticket sales trial as Helen Sharman spacesuit goes back on display

UK tour of Tim Peake's spacecraft attracts 1.3 million visitors as Science Museum marks Apollo anniversaries with Summer of Space



LOOK CLOSER

Helen Sharman on her Sokol space suit

Apply NER to Catalogue Descriptions…

British NORP astronaut, Helen Sharman's PERSON Sokol OBJECT spacesuit made by Zvezda ORG . Sharman PERSON wore this rescue suit during the space flight on board the SOYUZ-TM-12 and MIR spacecraft in May 1991 DATE . Space suit model number KV-2 No. 167 CARDINAL .

Sokol-KV-2 OBJECT rescue suit worn by Helen Sharman PERSON during the Juno OBJECT mission to the Mir OBJECT space station, 1991 DATE

СПАСАТЕЛЬНЫЙ СКАФАНДР

Helen Sharman PERSON was the first British NORP person in space. Sharman PERSON wore this suit for two hours on the ground to check its fit. Lying back, she tried to read but her arms ached from holding the book for so long. Despite the suit's cooling systems she sweated 2 litres during the mission launch. Once she could remove the suit, she dried it thoroughly to ensure it would not go mouldy.

The Sokol OBJECT suit was developed after three unsuited cosmonauts asphyxiated on the Soyuz 11 OBJECT mission in 1971 DATE when their descent module depressurised during the return to Earth LOC . Every cosmonaut now wears one during launch and return from space. It will keep the wearer alive for a number of hours in the event of a cabin depressurisation. Each suit is tailor made to the individual cosmonaut and comprises an inner, airtight 'bladder' of rubberised plastic and an outer layer of nylon canvas. There are connecting rings on the lower abdomen for air (cooling) and oxygen supplies and a centrally positioned pressure adjustment valve control on the chest; the pressure gauge is on the left wrist. The helmet and boots are integral with the rest of the suit; the gloves are attached with anodized aluminium bayonet fixings. Today DATE 's Sokol ORG design is little changed from the original.

…then train an entity linker to link entity mentions to records

Helen Sharman 1963

OCCUPATION: Astronaut, Broadcaster, Chemist, Engineer, Lecturer
NATIONALITY: British
BORN IN: Sheffield, South Yorkshire, England, United Kingdom

NPP Zvezda (Q541905)

company in Moscow, Russia edit
K-36DM | Zvezda (Russia) | Research-and-production enterprise "Zvezda" to them.
GI Severin | Zvezda Research and Production Enterprise

British NORP astronaut, Helen Sharman's PERSON Sokol OBJECT spacesuit made by Zvezda ORG . Sharman PERSON wore this rescue suit during the space flight on board the SOYUZ-TM-12 and MIR spacecraft in May 1991 DATE . Space suit model number KV-2 No. 167 CARDINAL .

Sokol-KV-2 OBJECT rescue suit worn by Helen Sharman PERSON during the Juno OBJECT mission to the Mir OBJECT space station, 1991 DATE

СПАСАТЕЛЬНЫЙ СКАФАНДР

Helen Sharman PERSON was the first British NORP person in space. Sharman PERSON wore this suit for two hours on the ground to check its fit. Lying back, she tried to read but her arms ached from holding the book for so long. Despite the suit's cooling systems she sweated 2 litres during the mission launch. Once she could remove the suit, she dried it thoroughly to ensure it would not go mouldy.

The Sokol OBJECT suit was developed after three unsuited cosmonauts asphyxiated on the Soyuz 11 OBJECT mission in 1971 DATE when their descent module depressurised during the return to Earth LOC . Every cosmonaut now wears one during launch and return from space. It will keep the wearer alive for a number of hours in the event of a cabin depressurisation. Each suit is tailor made to the individual cosmonaut and comprises an inner, airtight 'bladder' of rubberised plastic and an outer layer of nylon canvas. There are connecting rings on the lower abdomen for air (cooling) and oxygen supplies and a centrally positioned pressure adjustment valve control on the chest; the pressure gauge is on the left wrist. The helmet and boots are integral with the rest of the suit; the gloves are attached with anodized aluminium bayonet fixings. Today DATE 's Sokol ORG design is little changed from the original.

Sokol space suit (Q1197668)

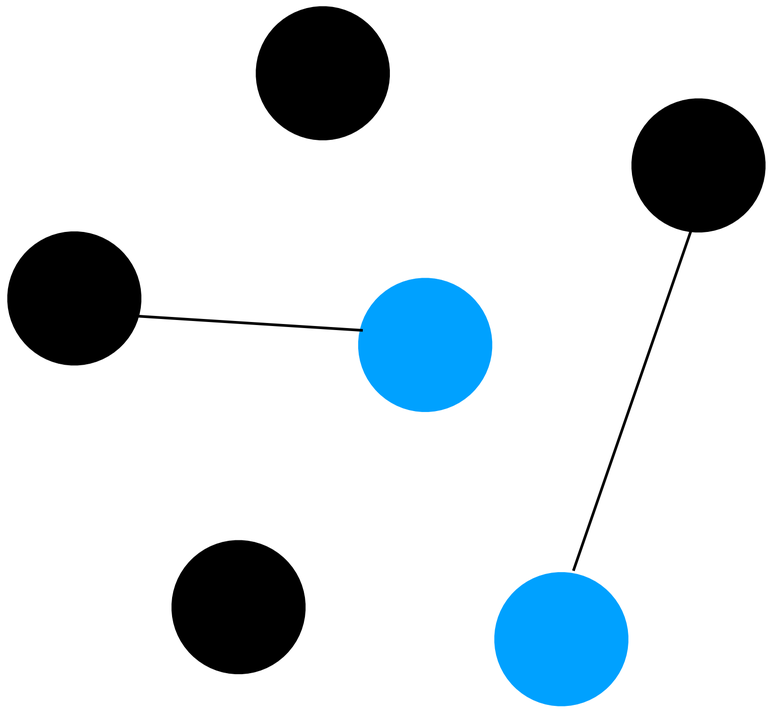
Russian spacesuit used on Soyuz
Sokol IVA | Sokol

Soyuz 11 (Q648581)

Manned Soviet space mission to the Salyut 1 Space Station

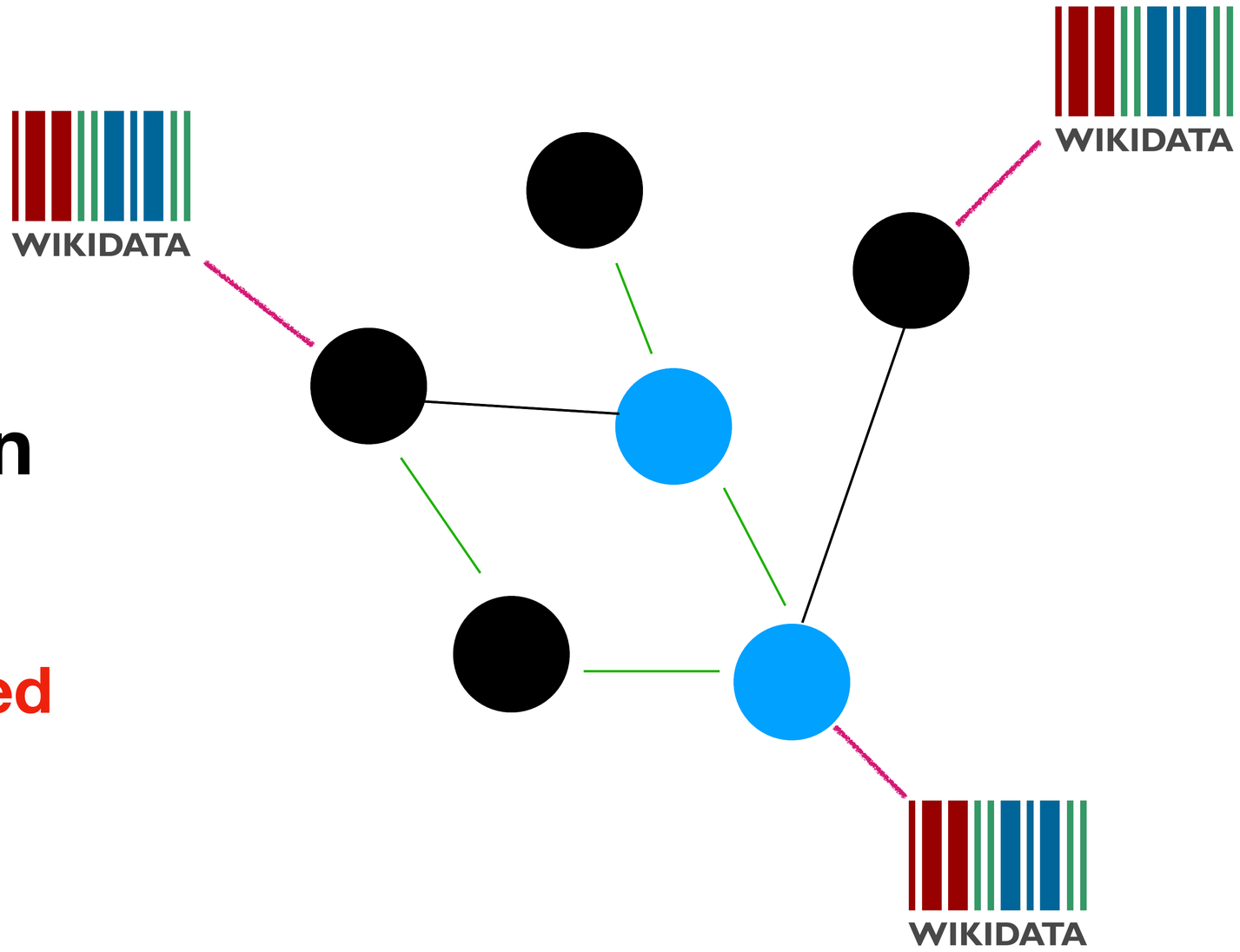
This was our collection before

Small islands of thin data



This is our collection interlinked via information extraction techniques

Small islands of connected and interlinked data



How can existing digital tools and methods be used to build relationships at scale between poorly and inconsistently catalogued digitised collection objects and other content sources?

Is such an approach scalable to larger volumes of content and different types of collections?

Where is the best use of human input in supporting such an approach? What expertise and skills are required for this input?

- OntoNotes corpus (English) ≃ 1.5M words.

- 59% of respondents from cultural heritage institutions said that a major hurdle to them adopting Wikidata IDs in their collection was time, resources, or the large amount of work required[1]

- What is the best use of human input → how can we quickly integrate human expertise into an NER model?

[1]Heritage Connector June 2020 Convening: https://thesciencemuseum.github.io/heritageconnector/events/2020/06/22/wikidata-and-cultural-heritage-collections-webinar/

Using Rules to Augment NER

Date detection ↑ 1.5%, collection & archive names ↑ 0.5%

```
DATE_PATTERNS = [
    {"label": "DATE", "pattern": [{"SHAPE": "dddd"}, {"ORTH": "–"}, {"SHAPE": "dddd"}]},  # 1984 – 1990 | 1984–1990
    {"label": "DATE", "pattern": [{"ORTH": "c."}, {"SHAPE": "dddd"}]},  # c. 1200
    {"label": "DATE", "pattern": [{"TEXT": {"REGEX": r"c\.\d{3,4}"}}]},  # c.1200
    {"label": "DATE", "pattern": [{"TEXT": {"REGEX": r"c\.\d{3,4}"}}, {"ORTH": "–"}, {"SHAPE": "dddd"}]},  # c.1200 – 1220 | c.1200–1220
    {"label": "DATE", "pattern": [{"TEXT": {"REGEX": r"\d{1,2}/\d{1,2}/(\d{4}|\d{2})"}}]},  # 03/12/2000
    {"label": "DATE", "pattern": [{"TEXT": {"REGEX": r"\d{1,2}\.\d{1,2}\.(\d{4}|\d{2})"}}]},  # 03.12.2000
    {"label": "DATE", "pattern": [{"SHAPE": "dd"}, {"ORTH": "–"}, {"SHAPE": "dd"}, {"ORTH": "–"}, {"SHAPE": "dddd"}]},  # 03–12–2000
    {"label": "DATE", "pattern": [{"SHAPE": "d"}, {"ORTH": "–"}, {"SHAPE": "dd"}, {"ORTH": "–"}, {"SHAPE": "dddd"}]},  # 3–12–2000
    {"label": "DATE", "pattern": [{"SHAPE": "dd"}, {"ORTH": "–"}, {"SHAPE": "d"}, {"ORTH": "–"}, {"SHAPE": "dddd"}]},  # 03–1–2000
    {"label": "DATE", "pattern": [{"SHAPE": "d"}, {"ORTH": "–"}, {"SHAPE": "d"}, {"ORTH": "–"}, {"SHAPE": "dddd"}]},  # 3–1–2000
    {"label": "DATE", "pattern": [{"SHAPE": "dddd"}, {"ORTH": "to"}, {"SHAPE": "dddd"}]},  # 1805 to 1860
]
```

```
COLLECTION_NAME_PATTERNS = [
    # TODO: use 'POS': 'PROPN' here instead of IS_TITLE: True for better detection of proper nouns
    {"label": "ORG", "pattern": [{'IS_TITLE': True, 'OP': '+'}, {'LOWER': 'collection'}]},  # Sforza collection
    {"label": "ORG", "pattern": [{'IS_TITLE': True, 'OP': '+'}, {'LOWER': 'archive'}]},  # Charles Urban archive
]
```

A Collection as a Dictionary

↑ 3.2%


```
{"label": "ORG", "pattern": "Thames Archway Company", "id": "https://collection.sciencemuseumgroup.org.uk/people/cp15926"}
{"label": "ORG", "pattern": "Hodbarrow Mining Company", "id": "https://collection.sciencemuseumgroup.org.uk/people/cp16807"}
{"label": "ORG", "pattern": "HMS Vanguard (1815)", "id": "https://collection.sciencemuseumgroup.org.uk/people/cp17108"}
{"label": "ORG", "pattern": "Wind Energy Group", "id": "https://collection.sciencemuseumgroup.org.uk/people/cp17473"}
{"label": "ORG", "pattern": "E R and F Turner Limited", "id": "https://collection.sciencemuseumgroup.org.uk/people/cp17945"}
{"label": "ORG", "pattern": "Baird Television Limited", "id": "https://collection.sciencemuseumgroup.org.uk/people/cp17663"}
{"label": "ORG", "pattern": "Alliance Box Company", "id": "https://collection.sciencemuseumgroup.org.uk/people/cp24886"}
{"label": "ORG", "pattern": "Hell", "id": "https://collection.sciencemuseumgroup.org.uk/people/cp21022"}
{"label": "ORG", "pattern": "Paradigm Models Limited", "id": "https://collection.sciencemuseumgroup.org.uk/people/cp22440"}
{"label": "ORG", "pattern": "City of York Council", "id": "https://collection.sciencemuseumgroup.org.uk/people/cp19207"}
{"label": "ORG", "pattern": "Kvaerner Masa-Yards", "id": "https://collection.sciencemuseumgroup.org.uk/people/cp24946"}
{"label": "ORG", "pattern": "Frederick Bateman and Company Limited", "id": "https://collection.sciencemuseumgroup.org.uk/people/cp20289"}
{"label": "ORG", "pattern": "Normal School of Science, Astronomy Laboratory", "id": "https://collection.sciencemuseumgroup.org.uk/people/cp20442"}
{"label": "ORG", "pattern": "T Green & Son Ltd", "id": "https://collection.sciencemuseumgroup.org.uk/people/cp20553"}
```

In a few lines of code

```
nlp = spacy.load(model_type)
nlp.add_pipe("date_matcher", before="ner")
nlp.add_pipe(
    "pattern_matcher",
    before="date_matcher",
    config={"patterns": constants.COLLECTION_NAME_PATTERNS},
)
nlp.add_pipe(
    "thesaurus_matcher",
    config={
        "case_sensitive": False,
        "overwrite_ents": False,
        "thesaurus_path": thesaurus_path,
    },
    after="ner",
)
```

kalyan.dutia@sciencemuseum.ac.uk

 @kdutia