

NIFify: Towards Better Quality Entity Linking Datasets [†]

Henry Rosales-Méndez, Aidan Hogan and Barbara Poblete

University of Chile

{hrosales, ahogan, bpoblete}@dcc.uchile.cl

May 14, 2019

[†] LA-WEB 2019 - 10th Latin American Web Congress

Example

Joseph Walter “Joe” Jackson, was a steelworker at U.S. Steel. In an interview with Martin Bashir for the 2003 documentary *Living with Michael Jackson*, Jackson recalled that Joe often sat in a chair with a belt in his hand as he and his siblings rehearsed.

Example - Entity Recognition

Joseph Walter “Joe” Jackson, was a steelworker at U.S. Steel. In an interview with Martin Bashir for the 2003 documentary Living with Michael Jackson, Jackson recalled that Joe often sat in a chair with a belt in his hand as he and his siblings rehearsed.

Example - Entity Disambiguation

Wikipedia, the free encyclopedia

Article Talk Read Edit View history Search Wikipedia

Martin Bashir

From Wikipedia, the free encyclopedia

For the former MSNBC show, see *Martin Bashir* (TV series).

Martin Henry Bashir (born 18 January 1960) is a British journalist.

After initially gaining prominence on British television with his BBC interview with Osama bin Laden, in 2003, he was controversially fly-on-the-wall documentary with pop singer Michael Jackson on the TV. Bashir was a political commentator for MSNBC, hosting *Martin Bashir 360* and *Conversations*.



Wikipedia, the free encyclopedia

Article Talk Read Edit View history Search Wikipedia

Joe Jackson (manager)

From Wikipedia, the free encyclopedia

Joseph Walter Jackson (born July 26, 1930) is an American talent manager, and is the patriarch of the Jackson family of entertainers which includes the pop stars Michael Jackson and Janet Jackson.

Contents [show]

Early life and ancestry

Jackson was born to Samuel Jackson (1880–1983), an university professor, and Crystal Lee King (1907–2007). In February 198, his parents, on July 26, 1930, born in 1929, according to the Jacksons' Black Hall of Fame and Jacksons' books fly family, the Jacksons' father was the oldest of five children, which included brother Lawrence (died 2007) and father and sister's wife and sister's wife. Jackson was the son of Isaac and Mary Jackson (1880–1983).



Wikipedia, the free encyclopedia

Article Talk Read Edit View history Search Wikipedia

U.S. Steel

From Wikipedia, the free encyclopedia (redirected from USS Steel)

The **United States Steel Corporation** (NYSE: *U.S.*), more commonly known as **U.S. Steel**, is an American integrated steel producer headquartered in Pittsburgh, Pennsylvania, with production operations in the United States, Canada, and Central Europe. As of 2015, the company was the world's 24th largest steel producer and second largest domestic.

United States Steel Corporation



Type	Public
Traded as	NYSE: <i>U.S.</i>
Industry	Steel

Joseph Walter "Joe" Jackson, was a steelworker at U.S. Steel. In an interview with Martin Bashir for the 2003 documentary *Living with Michael Jackson*, Jackson recalled that Joe often sat in a chair with a belt in his hand as he and his siblings rehearsed.

Wikipedia, the free encyclopedia

Article Talk Read Edit View history Search Wikipedia

Michael Jackson

From Wikipedia, the free encyclopedia

For other people named Michael Jackson, see *Michael Jackson (disambiguation)*.

"King of Pop" redirects here. For other uses, see *King of Pop (disambiguation)*.

Michael Joseph Jackson (August 29, 1958 – June 25, 2009) was an American singer, songwriter, and dancer. Dubbed the "King of Pop", he was one of the most popular entertainers in the world, and was the first selling music artist during the year of his death (19). Jackson's contributions to music, dance, and fashion (19) along with his publicized personal life made him a global figure in popular culture for most four decades.

The single child of the Jackson family, Michael made his professional debut in 1964 with his older brothers, Jackie, Tito, Jermaine, and Marlon as a member of the Jackson 5. He began his solo career in 1971 with the *Moonwalk* Records, in the early 1980s, Jackson became a global star in popular music, his music videos,



Wikipedia, the free encyclopedia

Article Talk Read Edit View history Search Wikipedia

Living with Michael Jackson

From Wikipedia, the free encyclopedia

Living with Michael Jackson is a television documentary, in which Martin Bashir interviewed Michael Jackson over a span of 8 months, from May 2002 to January 2003, about different aspects of his life. It was shown first in the United Kingdom on ITV (as a *Tonight special*) on February 3, 2003, and in the United States 3 days later on ABC, introduced by Bashir.

Living with Michael Jackson



Genre	Documentary
-------	-------------

Name Variations in Entity Linking

Michael Joseph Jackson

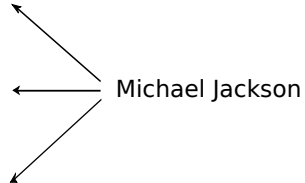
Michael J. Jackson

King of Pop



Name Variations in Entity Linking

The image displays three screenshots of Wikipedia articles, each showing a different name variation for the same entity, Michael Jackson. The top screenshot is the main article for "Michael Jackson", which includes a disambiguation note: "For other people named Michael Jackson, see Michael Jackson (disambiguation)." The middle screenshot is the article for "Michael Jackson (writer)", also with a disambiguation note. The bottom screenshot is the article for "Michael Jackson (actor)", which includes a warning box: "This biography of a living person needs additional citations for verification. Please help by adding reliable sources. Contentious material about living persons that is unsourced or poorly sourced must be removed immediately, especially if potentially libelous or harmful. (May 2016) (Learn how and when to remove this template message)." Each screenshot shows the Wikipedia logo, navigation links, and a search bar.



Michael Jackson

Are there benchmark datasets to measure Entity Linking results?

Overview of popular EL datasets

Dataset	Mn	Typ	Format
MSNBC	X	X	MSNBC
IITB	✓	X	IITB
AIDA/CoNLL	✓	X	AIDA
ACE2004	X	X	MSNBC
AQUAINT	X	X	MSNBC
DBpedia Spotlight	✓	X	Lexvo
KORE50	✓	X	AIDA
N3-RSS 500	✓	X	NIF
Reuters 128	✓	X	NIF
News-100	✓	X	NIF
Wes2015	✓	X	NIF
SemEval 2015 Task 13	✓	X	SemEval
Thibaudet	X	✓	RENDEN
Bergson	X	✓	RENDEN
DBpedia Abstracts	X	X	NIF
MEANTIME	✓	✓	CAT
VoxEL	✓	X	NIF

Overview of popular EL datasets

Dataset	Mn	Typ	Format
MSNBC	X	X	MSNBC
IITB	✓	X	IITB
AIDA/CoNLL	✓	X	AIDA
ACE2004	X	X	MSNBC
AQUAINT	X	X	MSNBC
DBpedia Spotlight	✓	X	NIF
KORE50	✓	X	NIF
N3-RSS 500	✓	X	NIF
Reuters 128	✓	X	NIF
News-100	✓	X	NIF
Wes2015	✓	X	NIF
SemEval 2015 Task 13	✓	X	SemEval
Thibaudet	X	✓	RENDEN
Bergson	X	✓	RENDEN
DBpedia Abstracts	X	X	NIF
MEANTIME	✓	✓	CAT
VoxEL	✓	X	NIF

Proposal

NIFify: a tool that simultaneously supports the creation, visualization, and validation of NIF datasets, as well as the comparison of EL systems.

Related Work

- NIF-Dataset creation

QRTool NIF-Dataset Creation



Demo



Source Code

BENGAL Automatic NIF Creation



Demo



Source Code

Related Work

- NIF-Dataset creation

QRTool NIF-Dataset Creation



Demo



Source Code

BENGAL Automatic NIF Creation



Demo



Source Code

- NIF-Dataset validation

Eaglet NIF-Dataset Validation



Demo



Source Code

Related Work

- NIF-Dataset creation

QRTool	NIF-Dataset Creation		 Demo		 Source Code
---------------	-----------------------------	--	--	--	---

BENGAL	Automatic NIF Creation		 Demo		 Source Code
---------------	-------------------------------	--	--	--	---

- NIF-Dataset validation

Eaglet	NIF-Dataset Validation		 Demo		 Source Code
---------------	-------------------------------	--	--	--	---

- Benchmarking

GERBIL	Benchmark	 Demo	 Source Code	 Visualization
---------------	------------------	--	---	--

Orbis	Benchmark	 Demo	 Source Code	 Visualization
--------------	------------------	--	---	--

Related Work

- NIF-Dataset creation

QRTool NIF-Dataset Creation | ❌ Demo | ❌ Source Code

BENGAL Automatic NIF Creation | ✅ Demo | ✅ Source Code

- NIF-Dataset validation

Eaglet NIF-Dataset Validation | ❌ Demo | ✅ Source Code

- Benchmarking


GERBIL Benchmark | ✅ Demo | ✅ Source Code | ❌ Visualization

Orbis Benchmark | ❌ Demo | ❌ Source Code | ✅ Visualization

- NIFify

NIFify NIF-Dataset Creation / Validation / Benchmarking
✅ Demo | ✅ Source Code | ✅ Visualization

NIFify - Creation

 NIFify tool














Document AnnotationACE FormatMSNBC FormatState-of-the-art SystemsValidationHelp

FilterClearRemove All


Enter the document URI

Joseph Walter "Joe" Jackson, was a steelworker at U.S. Steel.
In an interview with Martin Bashir for the 2003 documentary Living with Michael Jackson, Jackson recalled that Joe often sat in a chair with a belt in his hand as he and his siblings rehearsed.

Split in sentencesNew Document



#1



Joseph Walter "Joe" Jackson, was a steelworker at U.S. Steel.
In an interview with Martin Bashir for the 2003 documentary Living with Michael Jackson, Jackson recalled that Joe often sat in a chair with a belt in his hand as he and his siblings rehearsed.

Annotate EntityDelete Doc

1


Joseph Walter "Joe" Jackson, was a steelworker at U.S. Steel.

2

In an interview with Martin Bashir for the 2003 documentary Living with Michael Jackson, Jackson recalled that Joe often sat in a chair with a belt in his hand as he and his siblings rehearsed.



NIFify - Creation

 NIFify tool


Document AnnotationACE FormatMSNBC FormatState-of-the-art SystemsValidationHelp

FilterClearRemove All


Enter the document' URI

Joseph Walter "Joe" Jackson, was a steelworker at U.S. Steel.
In an interview with Martin Bashir for the 2003 documentary Living With Michael Jackson, Jackson recalled that Joe often sat in a chair with a belt in his hand as he and his siblings rehearsed.

Split in sentencesNew Document



#1



Joseph Walter "Joe" Jackson, was a steelworker at U.S. Steel.
In an interview with Martin Bashir for the 2003 documentary Living With Michael Jackson, Jackson recalled that Joe often sat in a chair with a belt in his hand as he and his siblings rehearsed.

Annotate EntityDelete Doc

1 Joseph Walter "Joe" Jackson was a steelworker at U.S. Steel.

2 In an interview with Martin Bashir for the 2003 documentary Living With Michael Jackson, Jackson recalled that Joe often sat in a chair with a belt in his hand as he and his siblings rehearsed.


NIFify - Creation

The screenshot displays the NIFify tool interface with a modal dialog box titled "Add Annotation" open. The background interface includes a header with the NIFify logo and navigation links, a document input field, and a text area containing a paragraph about Joseph Walter "Joe" Jackson. The dialog box contains the following elements:

- Add Annotation** (Title bar with a close button)
- Text:** "...t in a chair with a **belt** in his hand as he a..."
- List of URIs:** A text input field containing "https://en.wikipedia.org/wiki/Belt_(clothing)" and a dropdown menu set to "MISC" with a green "+ Add" button.
- Taxonomy:** A text input field containing "tax:Ambiguous".
- Comment:** A text area with the placeholder "Enter the comments here".
- Buttons:** "Annotate in this doc" (orange), "Annotate in all docs" (orange), "Close" (grey), and "Annotate!" (blue).
- Checkbox:** "Allow Overlaps" (unchecked).

The background interface also shows a sidebar with a document list, a top navigation bar with links like "Document Annotation", "APF Format", "MADIF Format", "State-of-the-art Systems", "Validation", and "Help", and a right sidebar with "Filter", "Clear", and "Remove All" options.

NIFify - Creation

 NIFify tool

Document AnnotationACE FormatMSNBC FormatState-of-the-art SystemsValidationHelp

FilterClearRemove All

Joseph Walter "Joe" Jackson, was a steelworker at U.S. Steel.
In an interview with Martin Bashir for the 2003 documentary Living With Michael Jackson, Jackson recalled that Joe often sat in a chair with a belt in his hand as he and his siblings rehearsed.

Split in sentencesNew Document

#1

Modify

Joseph Walter "Joe" Jackson, was a steelworker at U.S. Steel.
In an interview with Martin Bashir for the 2003 documentary Living With Michael Jackson, Jackson recalled that Joe often sat in a chair with a belt in his hand as he and his siblings rehearsed.

Annotate EntityDelete Doc

1

Joseph Walter "Joe" Jackson

, was a steelworker at

U.S. Steel

.

2

In an interview with

Martin Bashir

for the 2003 documentary

Living With

Michael Jackson

Jackson

recalled that

Joe


often sat in a chair with a

belt

in his hand as he and his

siblings rehearsed.

NIFify - Validation

 NIFify tool

Document Annotation ACE Format MSNBC Format State-of-the-art Systems Validation Help

Category Rule Validations

Current Validators







#	Name	Erros	Date	Time	Description	Action
✗ 1	Link validations	-	-	-	Here we check if each link is valid or not. Also, we identify if they are ambiguous or redirect Wikipages	<input checked="" type="button" value="🔍"/>
✗ 5	Spelling checker	-	-	-	Those entity mentions with special characters or spaces on the borders, or with missing characters are identified.	<input type="button" value="🔍"/>
✗ 6	Surface form checker	-	-	-	We check here if the label of the annotations match with their correspondig substrings.	<input type="button" value="🔍"/>



Showing













#	Message	Sentence
---	---------	----------



NIFify - Validation

#	Name	Erros	Date	Time	Description	Action
✓ 1	Link validations	3	28-04-2019	13:22:14	Here we check if each link is valid or not. Also, we identify if there are ambiguous or redirect Wikipages.	 
✓ 5	Spelling checker	1	28-04-2019	13:22:46	Those entity mentions with special characters or spaces on the borders, or with missing characters are identified.	 
✓ 6	Surface form checker	0	28-04-2019	13:22:30	We check here if the label of the annotations match with their corresponding substrings.	 

(3) Showing: Link validations  

#	Message	Sentence
2	The uri https://en.wikipedia.org/wiki/King_of_Pop of the mention Jackson is a redirect page.	2 In an interview with  for the 2003 documentary    recalled that  e often sat in a chair with a 
3	The uri https://en.wikipedia.org/wiki/Belt of the mention belt is a disambiguation page.	2 In an interview with  for the 2003 documentary    recalled that  e often sat in a chair with a 

Errors found in current NIF datasets

Dataset	Spelling Error	Link Error	Format Error
DBpedia Spotlight	8	23	4
N3-RSS 500	1	34	–
Reuters 128	4	71	–
News-100	9	1515	–
Wes2015	–	609	–
VoxEL	–	8	–

- https://users.dcc.uchile.cl/~hrosales/dataset_errors.html

Errors found in current NIF datasets

[DBpedia Spotlight](#)
[N3-RSS 500](#)
[Reuters 128](#)
[News-100](#)
[Wes2015](#)
[VoxEL](#)

DBpedia Spotlight

(8) Spelling Errors found

1

Mention *songs* is a substring of other word

41 In a sense it's an improvement on " **Rebirth** ," even if a couple of the so **ngs b** ear the **scars**

2


Mention *dance* is a substring of other word

43 A tall **dance** r feigns clumsiness with a slyly graceful gawkiness that would do **Buster Keaton** proud, stumbling around a **ladder** and a **chair** to the delight of the **audience**

3

Mention *man* is a substring of other word

44 **Sasha Waltz** & Guests (pictured), named for its **choreographer** , is the **house** **dance** ensemble at Radialsystem V (Holzmarktstr, 33; radialsystem.de), a **perform** **man** ce space that opened in **2006** in an old **brick** **Gothic** **pumping station** on the **Spree** River in **Berlin**


NIFify tool

Document Annotation
ACE Format
MSNBC Format
State-of-the-art Systems
Validation
Help

Joseph Walter "Joe" Jackson, was a steelworker at U.S. Steel.

In an interview with Martin Bashir for the 2003 documentary Living With Michael Jackson, Jackson recalled that Joe often sat in a chair with a belt in his hand as he and his siblings rehearsed.

Babely_only_named_entities
Babely_only_named_entities
Babely_with_concepts
DBpedia Spotlight
AIDA
FREME NER
FRED
Tagme2
FOX
StanfordEN-FOX
BalieEN-FOX
OpenNLPEN-FOX
IllinoisExtendedEN-FOX
Nerd-combined

https://example.org/doc2
Annotate Sent.
Annotate Doc.


Type of annotation: by Document
Annotator: Babely_only_named_entities

https://example.org/doc1

3 Joseph Walter "Joe" Jackson , was a steelworker at U.S. Steel .

4 In an interview with Martin Bashir for the 2003 documentary Living With Michael Jackson , Jackson recalled that Joe often sat in a chair with a belt in his hand as he and his siblings rehearsed.

NIFify - Benchmarking

 NIFy tool

Document Annotation
 ACE Format
 MSNBC Format
 State-of-the-art Systems
 Validation
 Help

Joseph Walter "Joe" Jackson, was a steelworker at U.S. Steel.








In an interview with Martin Bashir for the 2003 documentary Living With Michael Jackson, Jackson recalled that Joe often sat in a chair with a belt in his hand as he and his siblings rehearsed.

Babelify only named entities

https://example.org/doc1

Annotate Sent.

Annotate Doc.

microF1

macroF1

microF1: 0.62

microPrecision: 0.57

microRecall: 0.67

#0



Type of annotation: by Sentences

Annotator: FRED






https://example.org/doc1#char=0,255

Modify

1

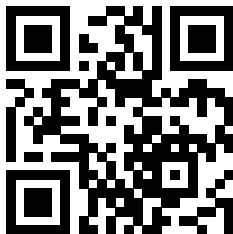
 Joseph Walter "Joe" Jackson
 , was a steelworker at
  U.S. Steel
 .

2

In an interview with
  Martin Bashir
 for the 2003 documentary
  Living With
  Michael Jackson
 ,
  Jackson
 recalled that
  Joe
 often sat in a chair with a belt in his hand as he and his siblings rehearsed.

Conclusion

- NIFify: Creation/Validation/Visualization/Benchmark
- Demo:
https://users.dcc.uchile.cl/~hrosales/NIFify_v2.html



- Source Code:
https://github.com/henryrosalesmendez/NIFify_v2

NIFify: Towards Better Quality Entity Linking Datasets [†]

Henry Rosales-Méndez, Aidan Hogan and Barbara Poblete

University of Chile

{hrosales, ahogan, bpoblete}@dcc.uchile.cl

May 14, 2019

[†] LA-WEB 2019 - 10th Latin American Web Congress