Umanistica Digitale - ISSN:2532-8816 - n.4, 2019
K. Eckert, M. Dadvar – JudaicaLink: A Knowledge Base for Jewish Culture and History
DOI: http://doi.org/10.6092/issn.2532-8816/9047

JudaicaLink: A Knowledge Base for Jewish Culture and History

[1]Kay Eckert and [2]Maral Dadvar

Web-based Information Systems and Services, Stuttgart Media University Germany
[1]eckert@hdm-stuttgart.de
[2]dadvar@hdm-stuttgart.de

Abstract. JudaicaLink is a novel resource which provides a knowledge base of Jewish culture and history. It is based on multilingual domain-specific information from encyclopedias and general-purpose knowledge bases such as The Integrated Authority File (GND) of the German National Library. JudaicaLink can be used for contextualization of metadata, i.e., entity resolution within and linking of metadata to improve resource access and to provide richer context to the user. Many resources for contextualization, particularly specialized resources for the given domain, are only available in unstructured form. General-purpose resources like DBpedia are hard to use due to their sheer size while only a very small subset of the data is actually relevant. Therefore, JudaicaLink aims at integrating relevant subsets of various data sources to function as a single hub for the contextualization process. JudaicaLink is freely available on the Web as Linked Open Data. In this paper, we explain how JudaicaLink is built, how it can be accessed by users, as well as its architecture, technical implementation, applications and relations to Jewish culture.

JudaicaLink è una nuova risorsa che fornisce una base di conoscenza sulla cultura e la storia ebraica. Si fonda su informazioni di dominio multilingua provenienti da enciclopedie e basi di conoscenza generiche come l'ntegrated Authority File (GND) della Biblioteca nazionale tedesca. JudaicaLink può essere utilizzato per contestualizzare i metadati, ovvero per nei processi di entity resolution e nel collegamento dei metadati, per fornire all'utente un migliore accesso alle risorse e un contesto informativo più ricco. Molte risorse per la contestualizzazione, in particolare risorse di dominio specializzate, sono disponibili solo in forma non strutturata. D'altra parte risorse generiche come quelle di DBpedia sono difficili da utilizzare, sia per via delle dimensioni sia perchè solo piccoli subset di dati risultano effettivamente rilevanti. Pertanto, JudaicaLink mira a integrare subset di dati rilevanti provenienti da varie fonti e svolgere in questo modo la funzione di un hub unico per il processo di contestualizzazione. JudaicaLink è liberamente disponibile sul Web in Linked Open Data. In questo paper viene illustrata la costruzione di JudaicaLink, le modalità di utilizzo da parte degli utenti, la sua architettura, l'implementazione tecnica, le sue applicazioni e relazioni con il patrimonio culturale ebraico.

Introduction

A knowledge base is a collection of knowledge about a variety of entities and it contains facts explaining those entities 1.. Besides being used for applications such as question answering 2., semantic search 3., visualization 4., and machine translation 5., knowledge bases also play an important role in information integration.

Some knowledge bases are specific to a certain domain such as occupations and job activities 6., others are general such as DBPedia[1] 7. and Yago[2] 1. which are huge sources of structured knowledge extracted from Wikipedia and other sources.

In this paper, we introduce JudaicaLink,[3] a new knowledge base specific to Jewish culture, history and studies. With JudaicaLink, we build a domain-specific knowledge base by extracting structured, multilingual knowledge from different sources (Figure 1).

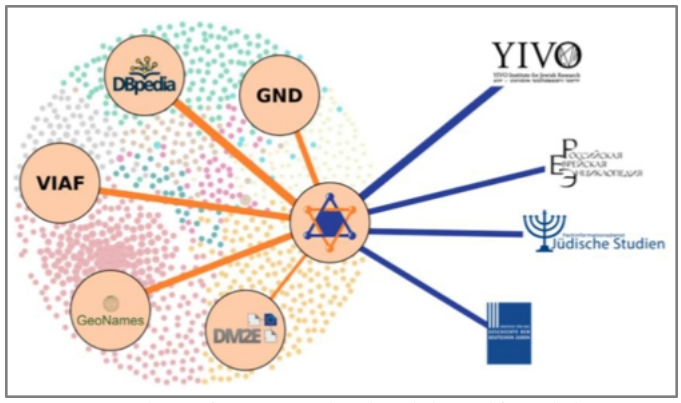

Figure 1: In JudicaLink, structured and multilingual knowledge is extracted from different sources and it functions as a central reference point for the contextualization of metadata

The main application of JudaicaLink so far is to act as a central reference point for the contextualization of metadata, *i.e.*, entity resolution within and linking of metadata to improve resource access and to provide richer context to the user. The task of contextualization consists of two steps; first, to identify entities unambiguously by means of stable URIs, e.g., a corresponding DBpedia resource or a GND identifier, and the second to find as much information (e.g., descriptions, links to related entities) as possible about the identified entity, usually by following links (like owl:sameAs) to other data sources and this way by obtaining further URIs suitable for identification. For an comprehensive example of JudaicaLink application for contextualization please refer to section Contextualization .

Example of data sources. Many useful data sources exist that can be used for contextualization

1 http://wiki.dbpedia.org/
2 http://yago-knowledge.org/
3 http://judaicalink.org/

in the domain of Jewish studies, e.g., domain-specific online encyclopedias like the YIVO Encyclopedia of Jews in Eastern Europe.[4] In contrast to general-purpose data sources such as DBpedia, this encyclopedia describes all entities in depth from a domain perspective, i.e., with respect to Jewish history. Such domain specific sources are more useful for scholars as they are trusted und citable publications while general-purpose data sources are much more interlinked and therefore helpful to create the graph structure.

As an example from YIVO encyclopedia (Figure 2), we can see that the available information about a topic, here the city of Minsk, is the title, describing text, the links in the text, a picture and the description of the picture. The links in the text points to other concepts and related material which can bring us more depth regarding a topic and can all be of great value for the domain studies.

On the other hand they lack the structured data access via Linked Open Data representations like DBpedia or Yago. Additionally, there are highly relevant data sources like the Integrated Authority File (GND) of the German National Library[5] providing mainly identifiers, which are used in libraries to contextualize literature, but also brief additional contextual information, usually of a very high quality. An unexpected drawback of these knowledge bases, however, is their sheer size. Setting up DBpedia or the GND for a local contextualization process is not a trivial task and requires considerable technical resources, despite the fact that only a very small part of these knowledge bases is relevant for the domain of Jewish studies.

Figure 2: Example of YIVO Encyclopedia

The challenges. In particular, there are three main problems that need to be dealt with; First, unstructured data sources like online encyclopedias need to be made available as structured data with stable URIs. Second, relevant subsets of general-purpose knowledge bases like DBpedia have to be identified to fill gaps between the specialized resources and to provide further context. And last, all data sources have to be integrated and interlinked.

4 http://yivoencyclopedia.org

5 http://dnb.de

The status. JudaicaLink is RDF-based and part of the Linked Open Data cloud. It includes information about persons, geographic places, subjects and occupations. At the time of this writing it contains 23,068 concepts, 17,739 persons and 460,102 triples. All data is available via URI dereferencing, a public SPARQL endpoint[6] and as data dumps.[7]

In the remainder of this paper we will first explain the data sources integrated into JudaicaLink so far and the infrastructure of this knowledge base. Then the applications and projects in which JudaicaLink has been used will be briefly introduced and finally we will address our next steps toward further developments of JudaicaLink.

Construction of Judaicalink

In this section we will describe the data sources which are integrated into JudaicaLink. We will explain the pros and cons of encyclopedias and general-purpose knowledge bases as data sources. Moreover, the infrastructure of the knowledge base, the data extraction process and representation are briefly explained.

Sources

Reference works such as encyclopedias and glossaries function as guides to specific scholarly domains. Therefore encyclopedias with a focus on Jewish studies were one of the sources of information in our knowledge base. The following encyclopedias have been so far integrated into JudaicaLink. What all these encyclopedias have in common is that they did not exist in a structured data format before. By using customized web scrapers, we extracted structured data from the article pages, e.g., the title, the article text, link relations to other articles.

Das Jüdische Hamburg

Das Jüdische Hamburg[8] (The Jewish Hamburg) is an encyclopedia containing articles in German by notable scholars about persons, locations and events of the history of Jewish communities in Hamburg. This is a free online resource based on the book "Das Jüdische Hamburg - Ein historisches Nachschlagewerk" 8.. It was published in 2006 on the occasion of the 40th anniversary of the Institute for the History of the German Jews.[9] It is a comparatively small dataset of 260 concepts. An excerpt of the extracted data is as follows:

```
<http://data.judaicalink.org/data/djh/ehre-ida> a foaf:Person;
jl:occupation <http://data.judaicalink.org/occupation/Schauspielerin> ;
jl:birthDate "9.7.1900" ;
jl:birthLocation "Prerau (Mähren)",
jl:deathDate "16.2.1989" ;
```

6 http://data.judaicalink.org/sparql
7 http://www.judaicalink.org/datasets
8 http://dasjuedischehamburg.de/
9 Institut für die Geschichte der deutschen Juden, IGdJ

```
jl:deathLocation "Hamburg",
jl: describedAt <http://dasjuedischehamburg.de/inhalt/ehre-ida> ;
jl:hasAbstract "Schauspielerin und Intendantin, geb. 9.7. 1900 Prerau
(Mähren), gest. 16.2.1989 Hamburg …" ,
skos:prefLabel "Ehre, Ida",
skos:altLabel "Ida Ehre".
```

Encyclopedia of Russian Jewry

Encyclopedia of Russian Jewry[10] provides an Internet version of the encyclopedia, which is published in Moscow since 1994, giving a comprehensive, objective picture of the life and activity of the Jews of Russia, the Soviet Union, and the Commonwealth of Independent States. The encyclopedia is structurally divided into three parts: 1) biographical information, 2) local history of the Jewish community in pre-revolutionary Russia, the Soviet Union and the CIS, and 3) thematic information on concepts related to Jewish civilization, the contribution of the Jews of Russia in various fields of activity, various Jewish social, scientific, cultural organizations, etc. The originally published volumes contain more than 10,000 biographies and more than 10,000 place names. The electronic version contains corrections and additions in the form of new articles, all in all 20,434 concepts. An excerpt of the extracted data is as follows:

```
<http://data.judaicalink.org/data/rdf/rujen/moskva> a skos:Concept;
       jl:describedAt<http://rujen.ru/index.php/%D0%BC%D0%BE%D1%81%D0%BA
%D0%B2%D0%B0> ;
       jl:hasAbstract "МОСКВА, город, центр Московской обл., столица
Российской Федерации. Изв. с 1147. В 14 в. – центр Вел. кн-ва Московского,
со 2-й пол. 15 в. – столица единого ...",
       jl:hasCategory <jld:rujen/geography> ;
       skos:related <jld:rujen/abelev_yuriy_matveevich>,
              <jld:rujen/abezgauz_evgeniy_zalmanovich>,
              <jld:rujen/abkin_abram_davydovich>,
              <jld:rujen/abragam_anatol>,
              <jld:rujen/abramov>,…
       owl:sameAs <http://data.dm2e.eu/data/place/onb/abo/Moscow>,
              <http://data.dm2e.eu/data/place/onb/abo/Moskau>,
              <http://data.dm2e.eu/data/place/sbb/kpe/Moskau>,
              <jld:yivo/Moscow>,
              <http://linkedgeodata.org/triplify/node1686293227>,…
       skos:prefLabel "Москва",
       skos:altLabel "Москва", "Москве", "Москвой", "Москву", "Москвы",… .
```

YIVO Encyclopedia

The YIVO Encyclopedia of Jews in Eastern Europe, courtesy of the YIVO Institute of Jewish Research, provides articles concerned with the history and culture of Jews in Eastern Europe from the beginnings of their settlement in the region to the present. The dataset contains 2,374 concepts. An excerpt of the extracted data is as follows:

```
<http://data.judaicalink.org/data/yivo/Moscow> a skos:Concept;
       jl:describedAt
<http://www.yivoencyclopedia.org/article.aspx/Moscow> ;
       jl:hasAbstract "Present-day capital of the Russian Federation, Moscow
```

10 http://rujen.ru/

```
was capital of the Russian state from ...",
       skos:related <jld:yivo/Agurskii_Samuil_Khaimovich> ,

              <jld:yivo/Akselrod_Zelik>, <jld:yivo/Altman_Natan_Isaevich>,
              <jld:yivo/America>,<jld:yivo/Anticosmopolitan_Campaign>,
              <jld:yivo/Antisemitic_Parties_and_Movements>, …
       owl:sameAs <http://data.dm2e.eu/data/place/onb/abo/Moscow>,
              <http://data.dm2e.eu/data/place/onb/abo/Moskau>,
              <http://data.dm2e.eu/data/place/onb/abo/Moskva>,
              <jld:rujen/Moskva>,
              <http://data.nytimes.com/moscow_russia_geo>,
              <http://dbpedia.org/resource/Moscow>, …
       skos:prefLabel "Moscow",
       skos:altLabel "Ida Ehre".
```

For extraction of the encyclopedias' contents we have made use of Coffeescript and Javascript modules. To this end, regular expression based methods were used for extraction of information, for example in Das Jüdische Hamburg. Encyclopedia, for extraction of birth date, death date birth location death location and occupation. Here we should emphasize on rich interlinking between the datasets. As illustrated in examples of Russian Jewry and YIVO encyclopedias, the common concepts, here "Moscow", are interlinked and connected as sameAs attributes.

There are also knowledge bases which contain a vast variety of information including facts related to Jewish culture. Therefore we also used these sources to extract a focused knowledge graph of concepts for the domain of Jewish studies.

DBPedia

DBPedia is a large-scale source of structured and multilingual knowledge extracted from Wikipedia. This knowledge base contains over 400 million facts that describe 3.7 million things 7.. We follow several approaches to extract relevant concepts from DBpedia: our main focus so far was on identifying prominent Jewish persons from different fields of activities. By identifying categories used to describe Jewish persons, we generated a list of these categories and searched for further persons. For each person, we extracted the name in all available languages, as well as links to other data sources. Typical categories include occupations, like "Rabbi".

As occupations are often available in other sources as well, we created an occupation ontology, combining labels and other information from various sources. The DBpedia dataset contains currently 5,294 persons with 35 distinct occupations. An excerpt of the extracted data is as follows:

```
<http://data.judaicalink.org/data/dbpedia/Aaron_Abiob> a foaf:Person;
       jl:occupation <http://data.judaicalink.org/data/occupation/Rabbi> ;
       dct:subject <http://dbpedia.org/resource/Category:16th-
century_rabbis> ;
       owl:sameAs <http://dbpedia.org/resource/Aaron_Abiob>,
                     <http://pt.dbpedia.org/resource/Aarão_Abiob>,
              <http://rdf.freebase.com/ns/m.02w_zfm>,
              <http://wikidata.dbpedia.org/resource/Q3624708>,
              <http://www.wikidata.org/entity/Q3624708>,
```

```
        <http://yago-knowledge.org/resource/Aaron_Abiob> ;
    skos:prefLabel "Aaron Abiob"@en, "Aarão Abiob"@pt.
```

GND

The Integrated Authority File (GND) of the German National Library is an authority file that contains identifiers to different concepts including to persons. Unlike DBPedia with its many categories, Jewish persons are not distinguished by any means in GND. Strategies to find relevant entries include the exploitation of publication data where the relevance can be determined via the publication. Occupations can also be used, but to a much smaller extent than in DBpedia, as DBpedia often contains specific categories for "Jewish authors", for instance, where GND only contains "author" as occupation. To narrow down our extraction, we assumed that writers and scholars from Israel are either Jew or their works are related to Jewish studies. Therefore, we also considered geographic information where available, for example for persons from Israel. For every person the name, occupation and identifiers were extracted. In the resulting RDF file the persons and their corresponding attributes were mapped to JudaicaLink ontology. This dataset includes 4,029 persons and 303 occupations. An excerpt of the extracted data is as follows:

```
<http://data.judaicalink.org/data/gnd/1022652842> a foaf:Person ;
        gndo:gndIdentifier <http://d-nb.info/gnd/1022652842> ;
        jl:occupation
<http://data.judaicalink.org/data/occupation/philosopher> ;
        skos:altLabel "Graupe, Heinz",
                "Graupe, Heinz M.",
                "Graupe, Heinz Moschke",
                "Graupe, Heinz-Moshe",
                "Heinz Mosche Graupe" ;
                skos:prefLabel "Graupe, Heinz Mosche" .
```

To extract the domain-specific graphs from the mentioned knowledge bases we used python code modules. All the extraction and data generation codes are available open source on our Github repository.[11] In the resulting RDF files the persons and their corresponding attributes were mapped to JudaicaLink ontology.

Infrastructure

JudaicaLink provides the datasets in N3 (Notation3)[12] and its subset formats, Turtle (Terse RDF Triple Language, TTL).[13] This format facilitates the usage and integration of JudaicaLink in triple stores and Semantic Web software such as Apache Jena.

The main JudaicaLink website is driven by the static site generator Hugo. We use the metadata of the web pages (Hugo frontmatter) (see Figure 3) to control the data publication process which is fully automated. On every push to the master branch, Github triggers an update script

11 https://github.com/wisslab/judaicalink-loader/
12 https://www.w3.org/TeamSubmission/n3/
13 https://w3.org/TR/turtle/

on our server that pulls the latest changes, rebuilds the website using Hugo and updates the data in the triple store according to the page metadata (see Figure 4).

```
+++
author = "Kai Eckert"
title = "Yivo Encyclopedia"
website = "http://www.yivoencyclopedia.org"
example = "http://data.judaicalink.org/data/yivo/Moscow"
graph = "http://data.judaicalink.org/data/yivo"
loaded = true
[[files]]
    url = "http://data.judaicalink.org/dumps/yivo/current/yivo.n3.gz"
    description = "Extraction from YIVO Encyclopediae"
+++

The YIVO Encyclopedia of Jews in Eastern Europe, courtesy of the YIVO Institute of
Jewish Research, NY.
<!--more-->

...
```

Figure 3: Dataset description in Markdown with metadata frontmatter

This way we ensure that the dataset descriptions on the web site, the data dumps and the data loaded in JudaicaLink are always consistent. The Hugo sources are collaboratively maintained using GitHub.

Every dataset corresponds to a name graph that can later on be accessed in the triple store. Datasets may consist of more than one data file since they might have been further expanded over time or may content different data components.

Users can download JudaicaLink datasets from the webpage of JudaicaLink.[14] The datasets can also be browsed as Linked Open Data using Pubby (with DM2E extensions) as Web Frontend 9.. Furthermore, a public SPARQL endpoint[15] is available.

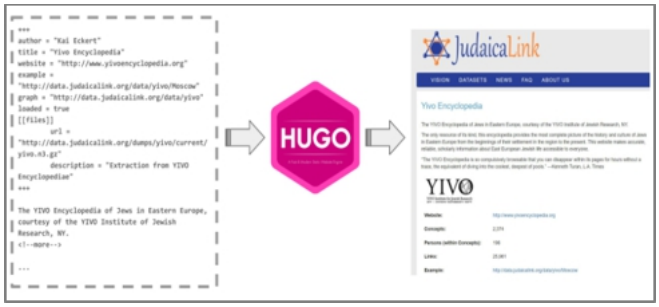

Figure 4: On every push to the master branch, the website using Hugo is rebuilt and according to page metadata the data in the triple store is updated.

14 http://www.judaicalink.org/datasets/
15 http://data.judaicalink.org/sparql

Ontology

The classes and properties used in JudaicaLink ontology are created on the fly based on the information that we encounter and need to be represented. However, we are consistent on the usage of the properties and the coined URI's are stable and unique. 5 illustrates the current state of the JudaicaLink ontology.

When a piece of information described in an encyclopedia is extracted, we assign the class 'Concept'. We use Natural Language Analysis techniques in order to analysis the concepts to identify whether they are a person. When identified as such, the class 'Person' is assigned to them and further properties are added. Every property of a Concept can be also used for a Person.

Application in the Specialized Information System for Jewish Studies Project

Contextualization

The Specialized Information System for Jewish Studies (SIS Jewish Studies) project[16] aims to create an expert information service for the domain of Jewish studies.

In this project which is developed at the University Library Johann Christian Senckenberg in Frankfurt, it is planned to produce a portal as a central platform for scientific information and among other purposes we aim to contextualize the extensive digital Judaica collections of the library. The entries of these collections include of titles along with their authors. Except for the name of the authors there is no other information available for each person. Therefore, we plan to enrich the authors' information with as much information as possible. For this purpose, based on the library database structure and other collections, every author is first enriched with their identifier in the Integrated Authority File of the German National Library, known as GND-id. This identifier would provide a wide range of information about each person.

JudaicaLink datasets are used as the source for this enrichment process. The corresponding id's were assigned to the authors based on the name of the authors as well as other features used for identity disambiguation, such as occupation and life span. So far 39.8% of the library digital collections have been contextualized using JudaicaLink. Two librarians have manually evaluated 10% of the contextualized records by and the contextualization accuracy was measured as 0.93.

Automatic Identification of Jewish Studies Titles

Beside contextualization, we also used JudaicaLink for automatic for automatic classification of Jewish studies titles 10.. There is a large number of titles in library collections which are not classified and indexed.

16 https://ub.uni-frankfurt.de/projekte/juedische_studien

NameSpaces	
NameSpace	**URI**
skos	http://www.w3.org/2004/02/skos/core
owl	http://www.w3.org/2002/07/owl
dc	http://purl.org/dc/terms/subject
gndo	http://d-nb.info/standards/elementset/gnd
jl	http://data.judaicalink.org/ontology
foaf	http://xmlns.com/foaf/0.1
dcterms	http://purl.org/dc/terms/identifier
rdfs	http://www.w3.org/2000/01/rdf-schema

Classes

Concept — **skos:Concept**

Name	Label	Namespace	URI	Description	Range
altLabel	alternative label	skos	skos:altLabel	Alternative writing format or language of the preferred label of a person or concept	xsd:string
prefLabel	preferred label	skos	skos:prefLabel	The label that is preferred to be used for a person or concept	xsd:string
related	has related	skos	skos:related	used to assert an associative link between two skos concepts	skos:Concept
scopeNote	scope note	skos	skos:scopeNote	Notes are used to provide information relating to skos concepts	xsd:string
broader	has broader	skos	skos:broader	used to indicates that one skos concept is in some way more general ("broader") than the other	skos:Concept
narrower	has narrower	skos	skos:narrower	used to indicates that one skos concept is in some way less general ("narrower") than the other	skos:Concept
primaryTopic	primary topic	foaf	foaf:primaryTopic	The primary topic of some page or document	owl:Thing
sameAs	same as	owl	owl:sameAs	These constructs may be used to create a number of different names that refer to the same individual	owl:Thing
identifier	identifier	dcterms	dcterms:identifier	An unambiguous reference to the resource within a given context	xsd:string
subject	subject	dc	dcterms:subject	The topic of the resource, category or subject heading that further describes this resource. (broader category used to describe)	owl:Thing
describedAt	described at	jl	jl:describedAt	Link to textual resource of encyclopedia article	foaf:Document
hasAbstract	has abstract	jl	jl:hasAbstract	The abstract of the article in which the concept or the person is described	xsd:string
hasCategory	has category	jl	jl:hasCategory	The category to which the concept was assigned to in the original data source	skos:Concept
referTo	refer to	jl	jl:referTo	Used for redirection of one concept to another concept mostly in encyclopedias (when there is a name change or other reasons)	skos:Concept

Person — **foaf:Person**

Name	Label	Namespace	URI	Description	Range
gndIdentifier	gnd identifier	gndo	gnd:gndIdentifier	The identifier of the person or concept in the Integrated Authority File of the German National Library (GND)	xsd:string
birthDate	birth date	jl	jl:birthDate	The date of birth of the person	xsd:date
deathDate	death date	jl	jl:deathDate	The date of death of the person	xsd:date
birthLocation	birth location	jl	jl:birthLocation	The location of birth of the person	xsd:string
deathLocation	death location	jl	jl:deathLocation	The location of death of the person	xsd:string
occupation	occupation	jl	jl:occupation	The occupation of the person	xsd:string

Document — **foaf:Document**

Name	Label	Namespace	URI	Description	Range
primaryTopic	primary topic	foaf	foaf:primaryTopic	The primary topic of some page or document	owl:Thing
label	label	rdfs	rdfs	used to provide a human-readable version of a resource's name	xsd:string
created	created	dcterms	dcterms:created	Date of creation of the resource	xsd:date

Figure 5: JudaicaLink ontology. This table represents classes Concept, Person and Document along with properties of each class.

These titles are spread and lost in a pool of data consisting of titles from many different fields such as chemistry, biology, linguistics, as well as Jewish studies. At University Library Johann Christian Senckenberg in Frankfurt a NLP (Natural Language Processing) based classification tool was developed to automatically identify the Jewish titles.

Our approach employs a classification model which processes the limited metadata available for each entry, using word level features and syntax level features. The word level features make use of a reference dataset, by which a wide range of prominent authors of Jewish literature can be identified. However, not any Jewish author was relevant to the classification. Therefore, another criteria which was set to determine whether an author is the relevant one or not, was to look them up in JudaicaLink and based on the other available information about the author in the knowledge base, decide whether they are the right person or not. A dataset of Jewish studies which was previously indexed was analyzed in order to extract the most informative elements of its titles. These elements were later on used by the syntax features to identify the Jewish studies in our non-indexed dataset. Before the actual analysis, a pre-processing step is involved, in order to remove the stopwords, data noise, and create the word stems.

Using our approach, we have analyzed 578'806 entries in the data pool, out of which 22'140 have been identified as Jewish studies. The performance accuracy was evaluated using an independent manually labeled dataset (n=18'872). The precision (positive predictive value) was 0.97, the recall (probability of detection) was 0.91, and F1 score (harmonic mean of precision and recall) was 0.94. The overall accuracy of the system was 89%. The tool was developed with reproducibility and adaptability in mind. The tool and the reference datasets are open source and accessible to be reused for similar purposes and in other libraries.

Conclusion and Future Work

In this paper we presented JudaicaLink, a knowledge base for Jewish literature and culture, which merges domain-specific information from different sources into one coherent entity. We described sources, the extraction process, and the applications of JudaicaLink. As future work, we would like to extend this project along different directions. We plan to extend our information extraction and textual analysis to the abstracts and definitions containing additional information. This information needs to be first identified then extracted and turned into triples.

So far we have used sources such as GND which contain more detailed information, however we will be expanding our data sources further to resources such as VIAF.[17]

The ontology used in our knowledge base will be improved and more specific as we progress and come across new concepts and entities. During our work we have come across researchers studying Holocaust who were interested in JudaicaLink and have found such knowledge base useful to their research. Therefore, although it is a bit different from the current topics in

17 http://viaf.org

JudaicaLink, we are looking into expanding our work to include Holocaust related information as well. The European Holocaust Research Infrastructure (EHRI) project[18] is working on controlled vocabularies related to Holocaust concepts, persons, corporate bodies and so on. As a step towards the exaptation of the topics of JudaicaLink, when the controlled vocabularies are published, they will be integrated into JudaicaLink. JudaicaLink is an ever growing source of information and adding new relevant resources is a continual goal towards making JudaicaLink a rich and comprehensive reference.

Acknowledgments

M. Dadvar is supported by FID Jüdische Studien (SIS Jewish Studies) project funded by German Research Foundation (DFG). We would like to acknowledge the support and contribution of Dov Winer from the European Association for Jewish Culture and Judaica Europeana and thank him for his valuable feedback and ideas. We would like to also thank Dr Rachel Heuberger from University Library of Frankfurt, and Marko Knepper from University Library of Mainz.

References

1. Rebele, Thomas, Fabian Suchanek, Johannes Hoffart, Joanna Biega, Erdal Kuzey, and Gerard Weikum. 2016. "YAGO: A Multilingual Knowledge Base from Wikipedia, Wordnet, and Geonames." In *The Semantic Web – ISWC 2016. Volume 9982 of LNCS.* Paul Groth et al. (eds.), 177–185. https://suchanek.name/work/publications/iswc-2016-yago.pdf
2. Kiyota, Yoji, Sadao Kurohashi, and Fuyuko Kido. 2002. "Dialog navigator: A question answering system based on large text knowledge base." *COLING 2002. Proceedings of the 19th international conference on Computational linguistics* 1, pp. 1–7. http://aclweb.org/anthology/C/C02/C02-1084.pdf
3. Tran, Thanh, Philipp Cimiano, Sebastian Rudolph, and Rudi Studer. 2007. "Ontology-Based Interpretation of Keywords for Semantic Search." *The Semantic Web, Volume 4825 of LNCS,* pp. 523-536. DOI: https://doi.org/10.1007/978-3-540-76298-0_38
4. Kraker, P., C. Kittel and A. Enkhbayar. 2016. "Open Knowledge Maps: Creating a Visual Interface to the World's Scientific Knowledge Based on Natural Language Processing." *027.7 Journal for Library Culture* 4/2, pp. 98–103. http://dx.doi.org/10.12685/027.7-4-2-157

18 https://www.ehri-project.eu

5. Knight, Kevin and Steve K. Luk. 1994. "Building a Large-Scale Knowledge Base for Machine Translation." *AAAI* 94, pp. 773–778. https://www.aaai.org/Papers/AAAI/1994/AAAI94-118.pdf

6. Gassen, Jonas B., Stefano Faralli, Simone P. Ponzetto and Jan Mendling. 2016. "Who-does-what: A knowledge base of people's occupations and job activities." *CEUR Workshop Proceedings* 1690, Paper 65. http://ceur-ws.org/Vol-1690/paper65.pdf

7. Lehmann, Jens, Robert Isele and Max Jakob. 2015. "Dbpedia – A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia." *Semantic Web Journal* 6/2, pp. 167–195. http://www.semantic-web-journal.net/system/files/swj558.pdf

8. Heinsohn, Kirsten. 2006. *Das juedische Hamburg: ein historisches Nachschlagewerk.* Göttingen:Wallstein Verlag.

9. Baierer, Konstantin et al.. 2017. "DM2E: A Linked Data Source of Digitised Manuscripts for the Digital Humanities." *Semantic Web Journal* 8/5, pp. 733–745. http://www.semantic-web-journal.net/system/files/swj1299.pdf

10. Dadvar, Maral, Rachel Heuberger, Annette Sasse and Kai Eckert. 2017. "Automatic NLP-based Classification of Jewish Studies Titles." *41st European Library Automation Group Conference (ELAG2017)*.