Quel che fa e quel che farà l'Intelligenza artificiale nei processi di pubblicazione dei dati bibliografici dell'iniziativa Share

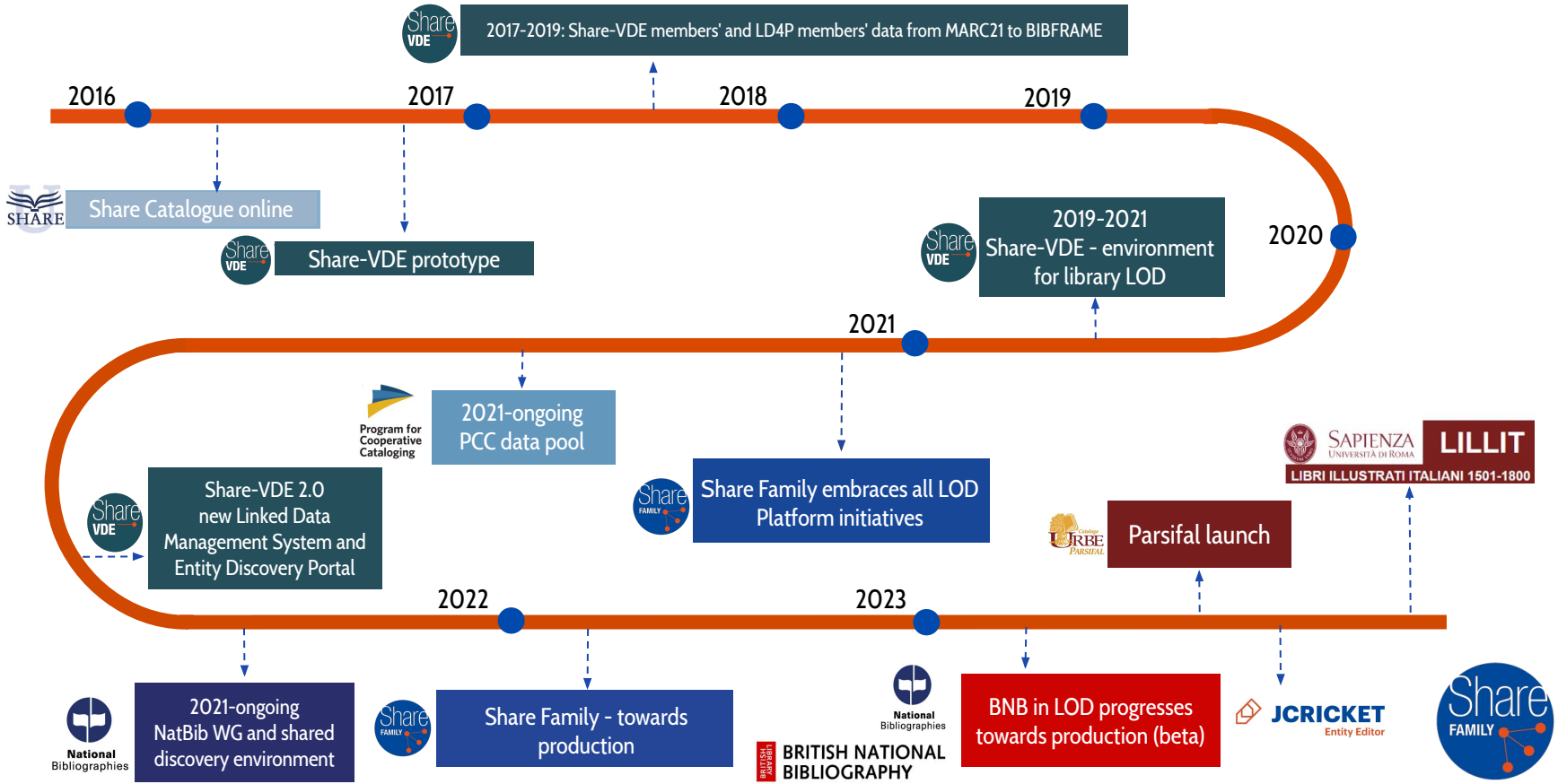
Tiziana Possemato & Andrea Gazzarini

62. Congresso Nazionale AIB
Biblioteche e tecnologie al tempo dell'Intelligenza Artificiale
Firenze - 16-17 novembre 2023

Agenda

- L'iniziativa Share Family: dove siamo
- Unsupervised Machine Learning: Clustering
- Supervised Machine Learning: Role Detection
- Supervised Machine Learning: Learning To Rank

La Share Family: un'iniziativa internazionale di condivisione



La Share Family: un ecosistema di Linked Open Data

La Share Family è una comunità globale basata sulla collaborazione tra biblioteche, archivi, musei, consorzi e sistemi gestionali (ILS/LSP) per condividere dati, servizi e competenze in un nuovo ecosistema bibliografico.

Per ulteriori dettagli fare riferimento a

https://www.share-family.org
e
the dedicated Share-VDE wiki section.

La Share Family: un ecosistema di Linked Open Data

INPUT DI DATI DALLE ISTITUZIONI

MARC21 bib. and holding

MARC21 aut.

UNIMARC

RDF/BIBFRAME

Other formats (eg. FOLIO)

X

LOD PLATFORM SERVICE

Mapping

URI Enrichment

Reconciliation

Creation of
Linked Data Entities

Conversion to
RDF/BIBFRAME

=

RISULTATI

[Data Publication]

- *End user discovery portals for each initiative of the Share Family*
- *Search APIs for data consumption (GraphQL, REST, SPARQL)*

+

[Data Distribution]

- *API / protocols for third parties integration (eg. local LSPs and data editor such as Wikidata, Sinopia BIBFRAME editor, FOLIO etc.)*
- *Downloadable datasets (BIBFRAME/RDF, MARC enriched records)*

+

[Shared Data Management]

- *Editing of Share Family entities with JCricket*
- *Reuse of Share Family BIBFRAME data in local library systems*
- *Reuse of Share Family MARC representations in local library systems*

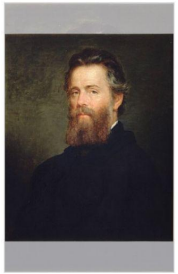
+

[Additional Services]

- *Authority control in MARC and BIBFRAME-based workflows*
- *SFI - Share Family Index: registry of entity URIs*

Navigare tra le entità nella Share Family

ShareVDE BETA | ← herman melville search | Search all content... | Options



Wikimedia Commons

Person

Herman Melville

American writer and poet (1819–1891). Born in New York (N.Y.). Died in New York (N.Y.).

Herman Melville was an American novelist, short story writer, and poet of the American Renaissance period. Among his best-known works are Moby-Dick (1851); Typee (1846), a romanticized account of his experiences in Polynesia; and Billy Budd, Sailor, a posthumously published novella. At the time of his death, Melville was no longer well known to the public, but the 1919 centennial of his birth was the starting point of a Melville revival. Moby-Dick eventually would be considered one of the great American novels. – Wikipedia

More options

Original works by | Publications by | Original works about

Moby Dick | Year | Creator | Genre

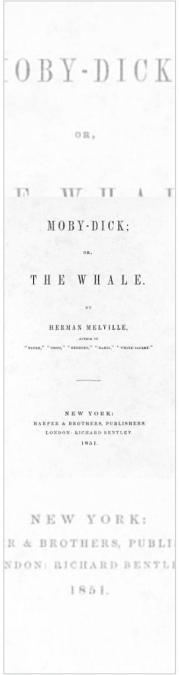
61 results

No.	Title	Year	Genre	Creators
1	Works. Selections	1987	Autobiographies/Fiction/Short stories/American literature	Herman Melville (author) A. Grove Day (other) Jay Leyda (other) Robert A. Sandberg (editor) Alma A. MacDougall (editor) Raymond M. Weaver (other) Harrison Hayford (other) Howard C. Horsford (other) Lynn Horth (other) G. Thomas Tanselle (editor)
2	Aspects of whiteness		Cantatas	John McCabe (author) Herman Melville (other)

Herman Melville in Share-VDE, con le sue opere, le pubblicazioni, le opere che parlano di lui

Navigare tra le entità nella Share Family

ShareVDE BETA | Search all content... | Options



Wikimedia Commons

Volume

Moby Dick

1851 novel by Herman Melville. Academic theses/biography/fiction/novel.

Moby-Dick; or, The Whale is an 1851 novel by American writer Herman Melville. The book is the sailor Ishmael's narrative of the maniacal quest of Ahab, captain of the whaling ship Pequod, for vengeance against Moby Dick, the giant white sperm whale that bit off his leg on the ship's previous voyage. A contribution to the literature of the American Renaissance, Moby-Dick was published to mixed reviews, was a commercial failure, and was out of print at the time of the author's death in 1891. Its reputation as a Great American Novel was established only in the 20th century, after the 1919 centennial of its author's birth. William Faulkner said he wished he had written the book himself, and D. H. Lawrence called it "one of the strangest and most wonderful books in the world" and "the greatest book of the sea ever written". Its opening sentence, "Call me Ishmael", is among world literature's most famous. – Wikipedia

This is part of the series Wright American fiction, Wright American fiction, 1851-1875, Digitalia e-books collection, Writings of Herman Melville, American Fiction, 1774-1920, Collectors' library, Wright American fiction. 1851-1875 . V. 2, First Avenue Classics (tm) Ser, Modern library, Verdensbiblioteket, Kunstforlagets illustrerte klassikere, Lommelydbok. Klassiker, Berømte klassikere fra Ex libris, Purŭn pada esŏ p'yŏlch'yŏ chinŭn kkum kwa pohŏm, Äventyrsklassiker, Vershyny svitovoho pysmenstva, Storie di mare, Shi jie wen xue ming zhu zhen cang ben, Great books of the Western world, First Avenue classics, Collection Folio classique, Riverside editions, Azbuka-klassika, Aschehoug pocket, Helikonbiblioteket, Barnes & Noble classics, Oxford world's classics (Oxford University Press), Project Gutenberg, Complete works, Ignatius critical editions, Library of America paperback classics, Grandi scrittori di ogni paese. Serie americana, Barnbiblioteket Saga, Akal bolsillo. Serie Clásicos de la literatura inglesa, Modern Library classics, Biblioteca Adelphi, Modern Library giants, World's classics, 100 greatest masterpieces of American literature, Library of literature, Modern library of the world's best books, Penguin classics, Longman critical edition, Wai guo wen xue ming zhu cong shu (Shanghai yi wen chu ban she), Akal bolsillo, Rinehart editions, Vershyny svitovoho pys'menstva, Vintage classics, Great books of the Western World (Franklin Center, Pa.), Works, Norton critical edition, Bibliothèque de la Pléiade, Oxford world's classics, Avant-scène. Théâtre, Armed Services editions (Series), Everyman's library. Fiction, 100 greatest books ever written, and Penguin classics deluxe edition.

More options

Publications of this | Related agents

L'opera Moby Dick in Share-VDE

Navigare tra le entità nella Share Family

ShareVDE BETA

Search all content...

"Modern Library classics" > "Moby Dick, or, The whale" (2000 edition in E...

Publication

Moby Dick, or, The whale

Volume. Published in English in 2000 in New York.

Contributors

Herman Melville (author), Barry Moser (artist), LeRoy Neiman (artist), Fred Benenson (editor), Sarah Shute (editor), Hershel Parker (editor), Harrison Hayford (editor), G. Thomas Tanselle (editor), Tony Tanner (editor), and Amazon Mechanical Turk (translator)

***ISBN**: 067978327X*
***Identifier**: p1241654873703306*
***Notes**: ill. ;*
***Bibliographic level**: Monograph*
***Cataloging source**: DLC*
***Credits**: Read by Paul Boehmer.*
***Descriptive cataloging form**: AACR*
***Dimension**: 21 cm.*
***Edition statement**: 2000 Modern Library pbk. ed.*
***Encoding level**: Full*
***Extent**: xl, 851 p. :*
***Record status**: Corrected or revised*
***Responsibility statement**: Herman Melville ; introduction by Elizabeth Hardwick ; illustrated by Rockwell Kent.*
***Series statement**: The Modern Library classics*
***System control number**: (OCoLC)ocm44541812*

Borrow a copy at UPenn Libraries | Publisher description | More options

Una delle numerose edizioni di Moby Dick, pubblicata a New York nel 2000.

Unsupervised Machine Learning: Clustering

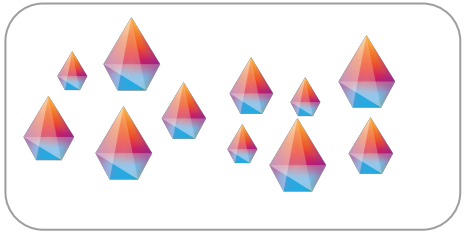
Clustering

Il concetto chiave del **clustering** è il **raggruppamento** di dati basato su **caratteristiche** condivise o simili.

Lo scopo è quello di **identificare**, all'interno di dataset corposI, complessi e/o eterogenei **strutture affini** tra loro **secondo una data prospettiva**.

Nella Share Family

In **Share-VDE** e in ognuno dei tenant appartenenti alla Share Family la fase di **clustering**, che costituisce la **prima parte** della **pipeline** di processo dei dati, ha lo scopo di **identificare**, **integrare**, **unire** e **arricchire** differenti **rappresentazioni** delle **stesse entità** provenienti dalle molteplici istituzioni che partecipano alla costruzione della base dati.



Clustering

Il Prisma

Un **tenant**, nella Share Family, è formato da un numero di **istituzioni** che **contribuiscono** alla **base di conoscenza** del progetto attraverso i loro dati, in svariati formati. Trattandosi di istituzioni di una certa rilevanza, e di un numero di contribuzioni consistente, è lecito aspettarsi una importante **percentuale di sovrapposizione** e **duplicazione** dei dati. I dati provenienti dai contributori contengono, in una certa misura, **differenti rappresentazioni** della **stessa entità**.

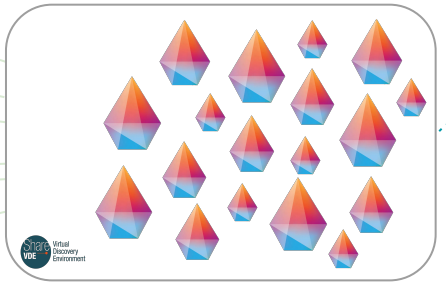
Il processo di clustering, in Share-VDE, produce delle **entità composite**, chiamate **Prisma**. Ciascun Prisma rappresenta **una medesima entità** nelle sue **variegate sfaccettature**.
Un Prisma è una entità che **raggruppa**, integra, **arricchisce** i dati di input omogenei rispetto a un concetto, e al tempo stesso mantiene i **dati originali provenienti** da ciascuna **contribuzione**.

Tenant e Provenance

Share-VDE gestisce una base di conoscenza formata da entità raggruppate, integrate e arricchite.

Un **tenant** è rappresentato da un insieme di istituzioni che contribuiscono alla stessa base di conoscenza.

Share-VDE Knowledge Base (**Sapientia**)

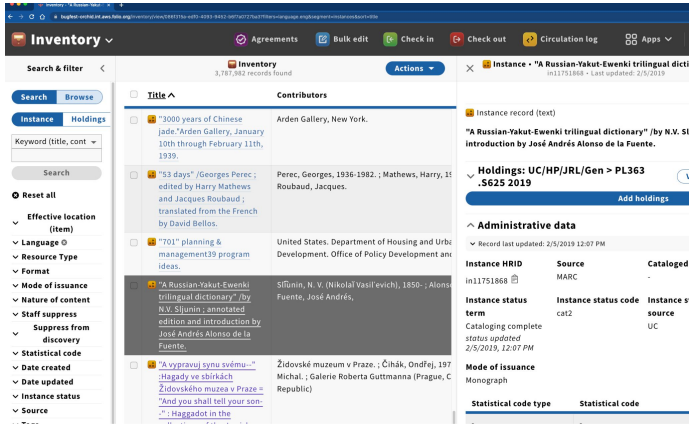


Una istituzione all'interno di un tenant viene definita come una **provenance**.
Al di là della terminologia, in Share-VDE viene sempre mantenuto un **legame** tra una **provenance** e i **dati che ha contribuito**.

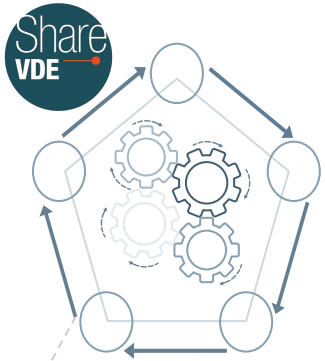
Dai dati della biblioteca alla Knowledge Base

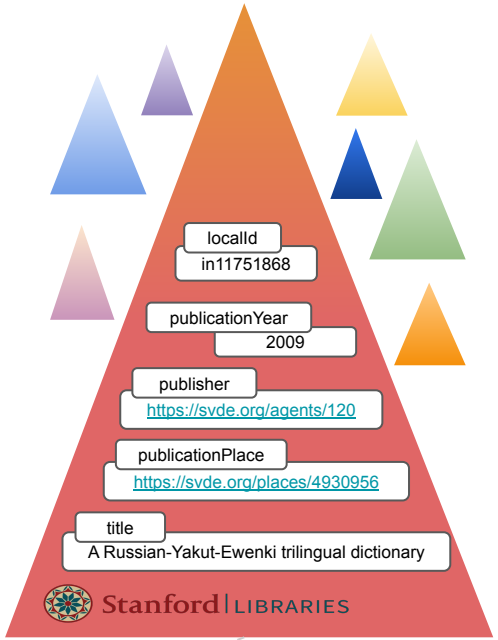
Una istituzione che partecipa a Share-VDE produce i suoi dati utilizzando FOLIO.

Stanford | LIBRARIES



I dati prodotti vengono inviati a Share-VDE

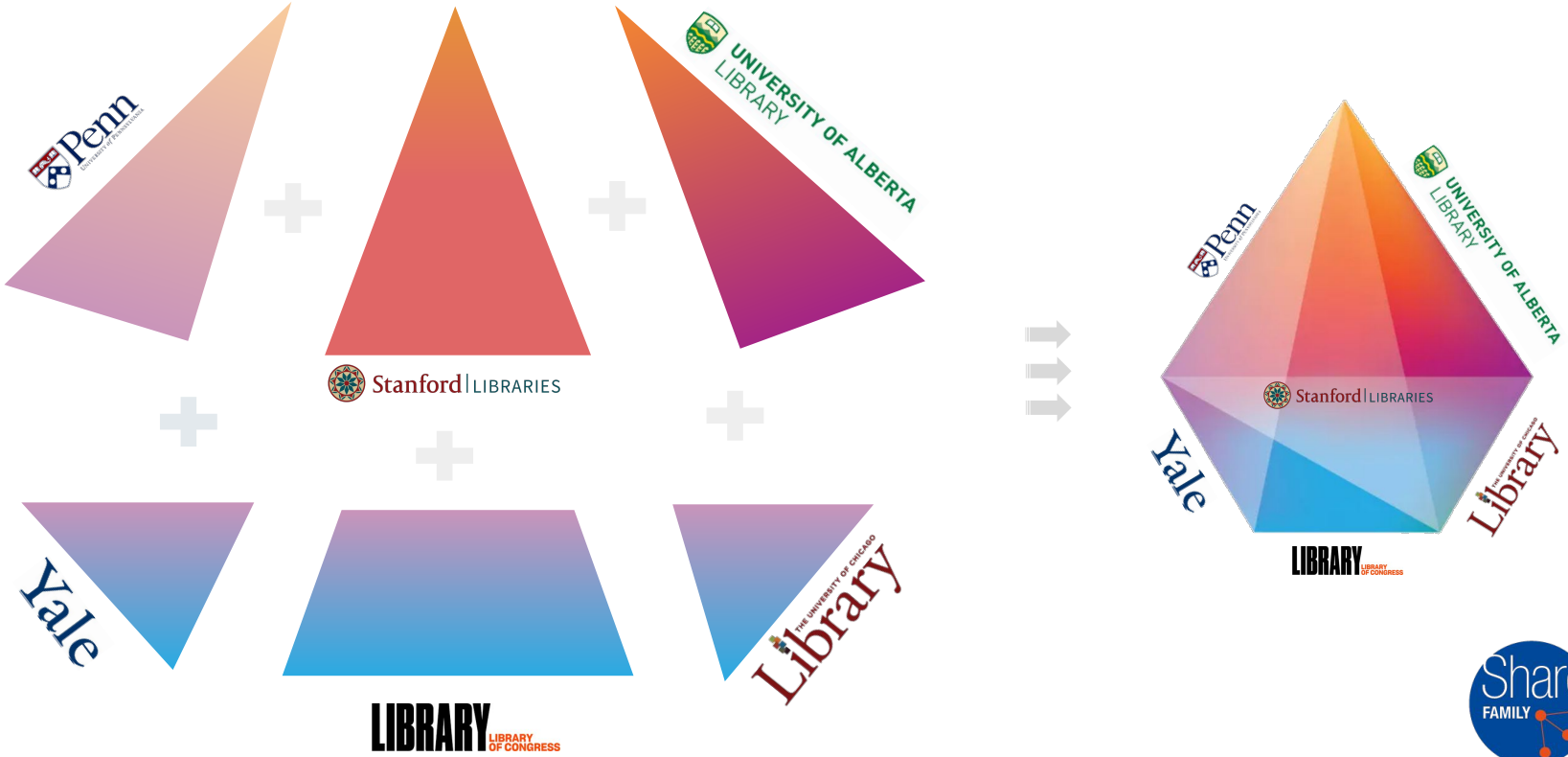




I dati vengono lavorati, divisi, raggruppati in riferimento e funzione alle entità che compongono il dominio di Share-VDE.

Share FAMILY

Il Prisma e le sue facce



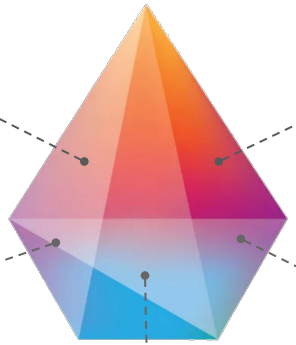
Il Prisma e le contribuzioni

UNIVERSITY OF ALBERTA LIBRARY

title	Alice in wonderland
titleAlternative	Alice's adventures under ground
author	https://svde.org/people/201

Stanford

title	Alice in wonderland
titleAlternative	Alice's adventures under ground
author	https://svde.org/people/201



LIBRARY LIBRARY OF CONGRESS

title	Alice in wonderland

National Library of Norway

title	Alice's adventures under ground
titleAlternative	Journeys in Wonderland
author	https://svde.org/people/201

(links)

sameAs	http://dbpedia.org/resource/Alice's_Adventures_in_Wonderland	Dbpedia
sameAs	https://www.wikidata.org/wiki/Q189875	Wikidata
sameAs	https://data.bnf.fr/ark:/12148/cb358500385#about	bnf

Attributi, Relazioni, Link

Name	Value	Provenance
title	Alice in wonderland	LIBRARY, Stanford
titleAlternative	Alice's adventures under ground	LIBRARY
titleAlternative	Journeys in Wonderland	National Library of Norway

Un attributo è una proprietà con un valore letterale.

sameAs	http://dbpedia.org/resource/Alice's_Adventures_in_Wonderland	Dbpedia
sameAs	https://www.wikidata.org/wiki/Q189875	Wikidata
sameAs	https://data.bnf.fr/ark:/12148/cb358500385#about	bnf

Un Link è una connessione tra una entità Share-VDE e una entità esterna.

Name	Provenance
author	LIBRARY, Stanford

Name	Value	Provenance
name	Carroll, Lewis	LIBRARY, Stanford
nameAlternative	Dodgson, Charles Lutwidge	LIBRARY, Stanford
nameAlternative	Karol, Luis	National Library of Norway

Una relazione è una connessione tra due entità Share-VDE.

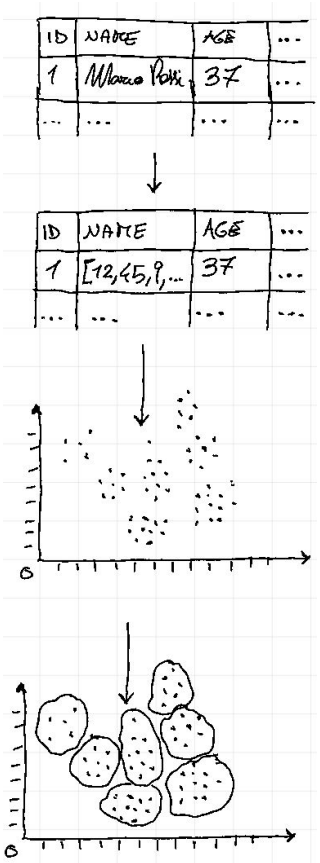
Clustering: Steps

Feature Engineering
Il dominio tematico di Share-VDE è composto da molteplici tipi di entità (e.g. agents, works, instances, items, places, formats, roles) ognuno dei quali è rappresentato da uno specifico insieme di **proprietà**. Tali proprietà, opportunamente selezionate e lavorate, diventato **caratteristiche** significative e rappresentative sulle quali basare la procedura di "raggruppamento" operata all'interno del clustering.

Una caratteristica, in questo contesto, è la misura o il valore di una grandezza riferita ad una data entità. L'output di questa fase è un dataset dove le entità sono rappresentate da un vettore (contenente elementi numerici e alfanumerici) di caratteristiche opportunamente lavorate.

Vettori numerici
Il vettore rappresentativo di ogni entità, come detto, è potenzialmente misto, contiene caratteristiche numeriche e alfanumeriche. In questa fase ogni caratteristica viene "tradotta" in un valore numerico.
Il vettore numerico che si ottiene rappresenta le coordinate dell'entità in uno spazio multidimensionale.

Clustering
La fase di clustering, utilizza, attraverso appositi algoritmi (unsupervised in questo caso) il dataset numerico output della fase precedente per trovare i cluster in accordo a diversi parametri forniti.



Supervised Machine Learning: Role Detection

La definizione dei ***ruoli*** nella strutturazione delle relazioni tra entità

ShareVDE BETA | Search all content... | Options

Wikimedia Commons

Volume

Blade runner 2049 (Motion picture)

2017 film directed by Denis Villeneuve.

Blade Runner 2049 is a 2017 American epic neo-noir science fiction film directed by Denis Villeneuve and written by Hampton Fancher and Michael Green. A sequel to the 1982 film Blade Runner, the film stars Ryan Gosling and Harrison Ford, with Ana de Armas, Sylvia Hoeks, Robin Wright, Mackenzie Davis, Dave Bautista, and Jared Leto in supporting roles. Ford and Edward James Olmos reprise their roles from the original film. Gosling plays K, a Nexus-9 replicant "blade runner" who uncovers a secret that threatens to destabilize society and the course of civilization. – Wikipedia

This is part of the series Philosophers on film.

More options

Publications of this | Related agents

Filter publications... | Publication year | Format | Publication place | Contributor | Type | Language

18 results

Sort by (A – Z)

1 Blade runner 2049 (Videodisc in English, published in 2018.)

Contributors:

- Philip K. Dick (creator)
- Benjamin Wallfisch (composer)
- Hans Zimmer (composer)
- Warner Bros. Home Entertainment (publisher)
- Warner Bros. Home Entertainment (Firm) (publisher)
- Bud Yorkin (film producer)
- Cynthia Sikes (film producer)
- Ridley Scott (film producer)

L'opera Blade runner 2049 (film)

La definizione dei ***ruoli*** nella strutturazione delle relazioni tra entità

ShareVDE BETA | Search all content... | Options

"Blade runner 2049 (Motion picture)" > "Blade runner 2049" (2018 edition in English)

Publication

Blade runner 2049

Videodisc. Published in English in 2018.

Contributors

Benjamin Wallfisch (composer), Hans Zimmer (composer), Ryan Gosling (actor), Harrison Ford (actor), Andrew A. Kosove (producer), Denis Villeneuve (director), Hampton Fancher (screenwriter), Michael Green (screenwriter), Roger A. Deakins (cinematographer), and Alcon Entertainment (Firm) (production company)

***Identifier**: p39041660062282195*
***Notes**: värillinen, ääni ;*
***Bibliographic level**: Monograph*
***Cataloging source**: FI-Piki*
***Credits**: Pääosissa: Ryan Gosling, Harrison Ford.*
***Descriptive cataloging form**: ISBD*
***Dimension**: 12 cm*
***Encoding level**: Core*
***Extent**: 1 Blu-ray-videolevy (2 h 43 min) :*
***Playback channel**: Surround*
***Record status**: Corrected or revised*
***Responsibility statement**: directed by Denis Villeneuve ; screenplay by Hampton Fancher and Michael Green ; produced by Andrew A. Kosove ... [ja muita].*
***System control number**: (FI-MELINDA)012213807*
***Video format**: Blu-Ray video*
***Video recording number**: 52GSB3100461*

Borrow a copy at NLF | More options

Una delle numerose pubblicazioni di Blade runner 2049 (2018) con tutti gli agenti ad essa collegati

Related agents | Subjects

Filter agents... | Start year | Location | Type

10 results

Sort by (A – Z)

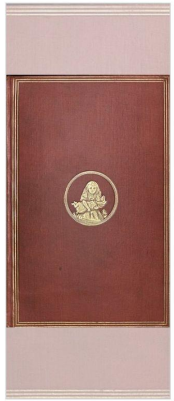
1. Benjamin Wallfisch (Person, Born in 1979.)
   Location: Born in London (England).
   Relationship: composer
2. Hans Zimmer (Person, Born in 1957.)
   Location: Born in Frankfurt am Main (Germany).
   Relationship: composer
3. Ryan Gosling (Person, Born in 1980.)
   Location: Born in London (Ont.).
   Relationship: actor
4. Harrison Ford (Person, Born in 1942.)
   Location: Born in Chicago (Ill.).
   Relationship: actor
5. Andrew A. Kosove (Person)
   Location: ∅
   Relationship: producer
6. Denis Villeneuve (Person, Born in 1967.)
   Location: Born in Bécancour (Québec).
   Relationship: director
7. Hampton Fancher (Person, Born in 1938.)
   Location: Born in Los Angeles (Calif.).
   Relationship: screenwriter
8. Michael Green (Person)
   Location: ∅
   Relationship: screenwriter
9. Roger A. Deakins (Person, Born in 1949.)
   Location: Born in Torquay (England).

Quando i metadati descrittivi sono parziali o insufficienti

ShareVDE BETA

Search all content...

Options



Wikimedia Commons

Volume

Alice's adventures in Wonderland

Novel by Lewis Carroll. 1865 dictionaries/fiction/short stories/children's stories.

Alice's Adventures in Wonderland is an 1865 English children's novel by Lewis Carroll, a mathematics don at Oxford University. It details the story of a young girl named Alice who falls through a rabbit hole into a fantasy world of anthropomorphic creatures. It is seen as an example of the literary nonsense genre. The artist John Tenniel provided 42 wood-engraved illustrations for the book. – Wikipedia

This is part of the series Nineteenth Century Collections Online: Children's Literature and Childhood, Naxos Music Library, First Avenue Classics (tm) Ser, Berømte bøker, Rainbow classics, Altemus' young people's library, Livres roses pour la jeunesse, Library for young people, Classics reimagined series, Readers library, Biblioteca pedagógica brasileira. Literatura infantil. Série 1a., Paragon children's library, Boeiende boek, Educator classic library, American Institute of Graphic Arts Keepsake, Broadview literary texts, Oxford world's classics (Oxford University Press), Colección Historias. Selección, Everyman's library children's classics, Riccardi Press books, Modern Library classics, Signet classic, Simon and Schuster classics, Golden classic, Sprīdīša bibliotēka, Broadview editions, Oscar narrativa, Insel Taschenbuch, DTV Junior, and Canterbury classics.

More options

Publications of this | Related agents

Filter publications... | Publication year | Format | Publication place | Contributor | Type | Language

245 results

Sort by (A – Z)

1 Ājaba deśe Ayālisa (Volume in Bengali, published in 2016.)

Contributors:

- Lewis Carroll (author)
- Rājaśrī Mukhopādhyāya (translator)
- Sumitā Sāmanta (translator)
- Bai-Citra Prakāśana (publisher)

L'opera Alice's adventures in Wonderland di Lewis Carroll, con le sue numerosissime edizioni

Quando i metadati descrittivi sono parziali o insufficienti

"Alice's adventures in Wonderland" › "Ājaba deśe Ayālisa" (2016 edition in Bengali)

Publication

Ājaba deśe Ayālisa

Volume. Published in Bengali in 2016 in Kalakātā.

Contributors

Lewis Carroll (author), Rājaśrī Mukhopādhyāẏa (translator), Sumitā Sāmanta (translator), and Bai-Citra Prakāśana (publisher)

***ISBN**: 9788193232897*
***Identifier**: p1041654873976698*
***Notes**: illustrations (black and white) ;*
***Bibliographic level**: Monograph*
***Cataloging source**: IiNdDKA*
***Descriptive cataloging form**: ISBD*
***Dimension**: 23 cm +*
***Encoding level**: Full*
***Extent**: 200 pages :*
***Parallel title**: Ajab deshe Alice*
***Record status**: New*
***Responsibility statement**: Liuisa Kyārala ; anubāda-ṭīkābhāshya, Rājaśrī Mukhopādhyāẏa-Sumitā Sāmanta = Ajab deshe Alice / b Carroll ; translation in Bengali by Rajasri Mukhopadhyay, Sumita Samanta.*
***System control number**: (OCoLC)1097563593*

Borrow a copy at UPenn Libraries

More options

Un'edizione in lingua bengali di Alice's adventures in Wonderland, dove la definizione dei ruoli è espressa a monte nel record bibliografico

Borrow a copy at UPenn Libraries

More options

Related agents | Subjects

Filter agents... | Location | Type

4 results

No.	Name	Type	Start year	End year	Location	Relationship
1	Lewis Carroll	Person	1832	1898	Born in Daresbury (England). Died in Guildford (England).	author
2	Rājaśrī Mukhopādhyāẏa	Person				translator
3	Sumitā Sāmanta	Person				translator
4	Bai-Citra Prakāśana	Organization				publisher

Quando i metadati descrittivi sono parziali o insufficienti

ShareVDE BETA | Search all content... | Options

"Alice's adventures in Wonderland" › "Ania v stranie chudes" (1982 edition in Russi...

Publication

Ania v stranie chudes

Volume. Published in Russian in 1982 in Ann Arbor.

Contributors

Lewis Carroll (author), M. S. Blekhman (translator), Alliance de recherche sur les discriminations (publisher), D. M. Urnov, Vladimir Vladimirovich Nabokov, S. A. Zalshupin, W. H. Auden, and N. M. Demurova

***Identifier**: p1071654874060372*
***Notes**: ill. ;*
***Bibliographic level**: Monograph*
***Cataloging source**: NjP*
***Descriptive cataloging form**: AACR*
***Dimension**: 22 cm.*
***Encoding level**: Full*
***Extent**: 114 p. :*
***Record status**: Corrected or revised*
***Responsibility statement**: L. Karrol' ; perevod s angl. V. Nabokova s risunkami Dzhona Tennielia.*
***System control number**: (OCoLC)ocm08714043*

Borrow a copy at UPenn Libraries | More options

Related agents | Subjects

Filter agents... | Start year | End year | Location | Type

Un'edizione in russo di Alice's adventures in Wonderland, dove la definizione dei ruoli non è espressa a monte nel record bibliografico

Related agents | Subjects

Filter agents... | Start year | End year | Location | Type

8 results

No.	Name	Type	Start year	End year	Location	Relationship
1	Lewis Carroll	Person	1832	1898	Born in Daresbury (England). Died in Guildford (England).	author
2	M. S. Blekhman	Person			Born in Ukraine.	translator
3	Alliance de recherche sur les discriminations	Organization				publisher
4	D. M. Urnov	Person	1936		Born in Moscow (Russia).	other
5	Vladimir Vladimirovich Nabokov	Person	1899	1977	Born in Saint Petersburg (Russia). Died in Montreux (Switzerland).	other
6	S. A. Zalshupin	Person	1900		Born in Saint Petersburg (Russia). Died in Paris (France).	other
7	W. H. Auden	Person	1907	1973		other
8	N. M. Demurova	Person	1930			other

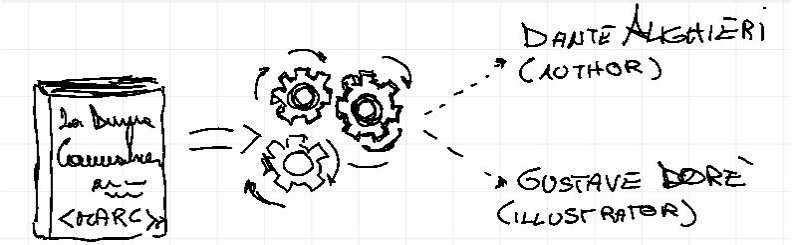
End of list

Role Detection

Un **record bibliografico** è una **rappresentazione strutturata** di un **documento bibliografico**, che può essere un libro, un articolo, una tesi, una registrazione audio o qualsiasi altro tipo di materiale informativo. Questa descrizione contiene **informazioni** che aiutano a **identificare** e **recuperare** il **documento** in un sistema di **catalogazione** o in una **biblioteca**.

L'obiettivo di questo ambito di ricerca e sviluppo è di **determinare**, a partire da un record bibliografico, i **nomi** che intervengono nella sua definizione in qualità di **contribuzioni**, assieme alla **qualifica** di tali contribuzioni.

Le contribuzioni all'interno di un record infatti possono essere espresse in forma **esplicita** o **strutturata**, oppure **implicita** o non **strutturata**. Ciò rende particolarmente **difficile** l'applicazione di una **logica procedurale** e **deterministica** per arrivare all'output desiderato. In aggiunta, le **informazioni** utili e **rilevanti** alla determinazione delle coppie **nome-ruolo** possono essere espresse con **abbreviazioni**, **codici** e in **molteplici lingue**.



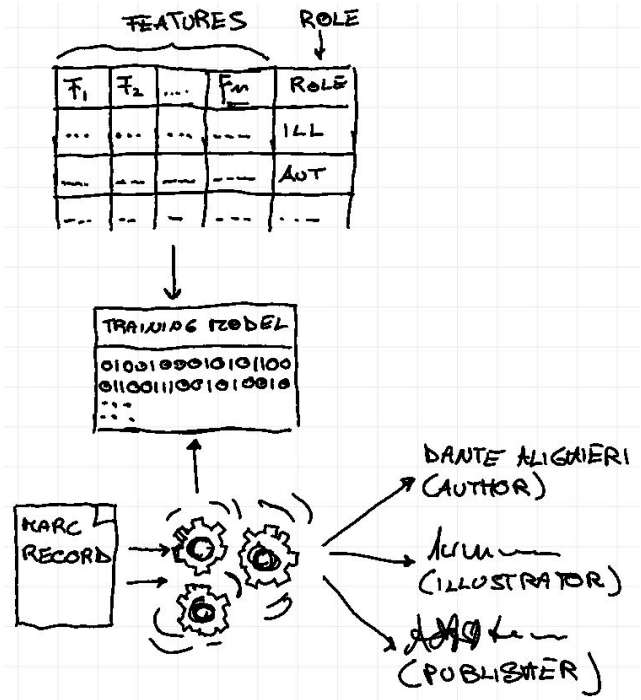
Role Detection: Steps

Feature Engineering
La fase è assolutamente identica, da un punto di vista formale, a quella vista nelle slide precedenti per il clustering. Chiaramente, le **features rilevanti** in questo contesto sono differenti.

L'obiettivo è **costruire**, più o meno in maniera **automatica**, un **dataset "controllato"** che verrà utilizzato come **base per l'addestramento**. In questo dataset, ogni **riga** rappresenta un nome all'interno di un record bibliografico con le **sue features** e il **ruolo** che il nome riveste in quel contesto.

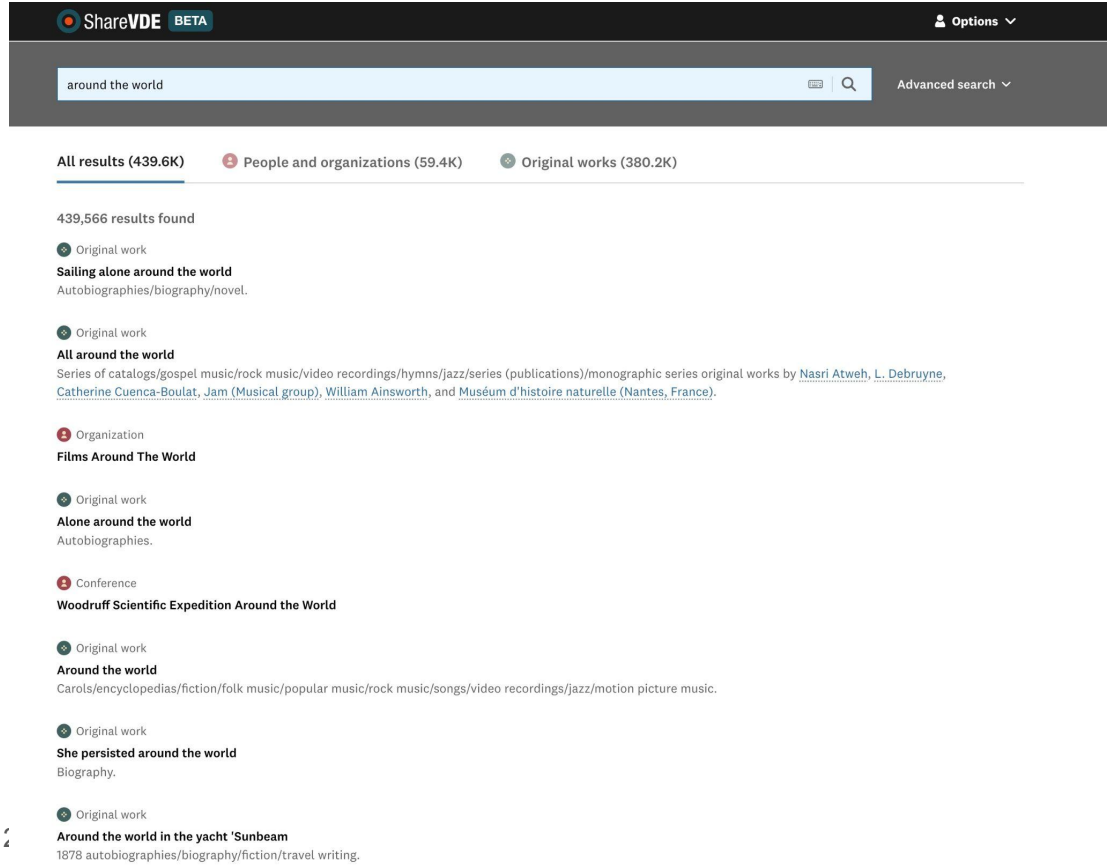
Addestramento
Il dataset ottenuto viene utilizzato per creare un **modello di apprendimento**, che guiderà il processo di **classificazione**.

Classificazione
Utilizzando il **modello di addestramento** fornito, il sistema, il cui **input** in questa fase è un **record bibliografico**, **classifica** i **nomi** presenti in base al **ruolo** desunto. Si parla di **classificazione** in quanto il **numero di ruoli** è definito a **priori** e pertanto lo **scopo** consiste **nell'individuare** la **classe** (i.e., il ruolo) **di appartenenza** di un determinato **nome**.



Supervised Machine Learning: Learning To Rank

Ottimizzare la *rilevanza* nei risultati di ricerca



Learning To Rank

Il **Learning to Rank** (LTR) è una tecnica di **Machine Learning** (supervisionata, in questo caso) utilizzata nel campo dell'**Information Retrieval** (IR) per **migliorare** ed **ottimizzare** la **fruizione dei risultati di ricerca** dal punto di vista della **rilevanza**, per l'**utente finale**.

Lo strumento primario di un sistema LTR è un **modello di apprendimento** che viene successivamente utilizzato da un **motore di ricerca** per **manipolare** e **definire** la **rilevanza** e di conseguenza **l'ordinamento** dei **risultati di ricerca** presentati all'utente.

Il modello viene creato utilizzando una **quantità considerevole** di **dati**, **segnali**, **interazioni** provenienti dall'utilizzo del sistema da parte degli utenti stessi.

L'applicazione del Learning To Rank in **contesti bibliografici** offre interessanti nonché stimolanti spunti di riflessione: esso viene infatti solitamente utilizzato in **contesti commerciali** (e.g. e-commerce) dove esistono **segnali** di **rilevanza espliciti** (e.g. "aggiungi al carrello", "acquista"). Tali segnali sono invece **solitamente assenti** o comunque **molto rari** in un **dominio bibliografico**.

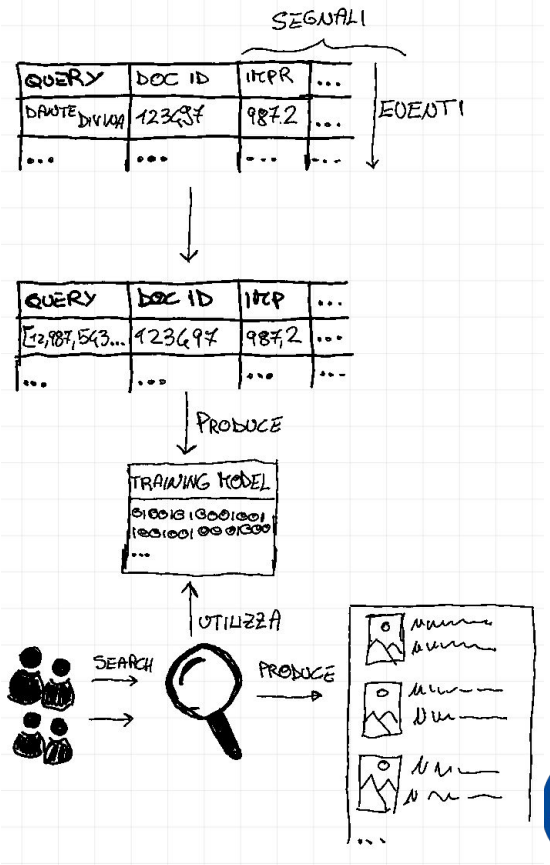
Learning To Rank: Steps

User Interaction Collection
Ogni interazione dell'utente con il sistema/portale viene collezionata e memorizzata insieme alla ricerca cui l'interazione risulta collegata. Le interazioni sono pertanto eventi che si verificano a partire da una ricerca iniziale, ed includono impressions, click, click secondari, scorrimenti e tempo di attesa su una pagina.

Feature Engineering
La fase è assolutamente identica, da un punto di vista formale, a quella vista nelle slide precedenti per il clustering. Chiaramente, le features rilevanti in questo contesto sono differenti.

Addestramento
Il dataset ottenuto viene utilizzato per creare un modello di apprendimento, che guiderà il re-ranking dei risultati in fase di ricerca.

Ricerca e (Re)Rank
Gli utenti utilizzano il sistema eseguendo delle ricerche; il motore applica le sue logiche interne per effettuare tali ricerche, e una volta ottenuta la lista dei risultati, utilizza il modello di apprendimento creato precedentemente per "riordinare" la lista in accordo alle regole impartite dal modello.



Grazie

Tiziana Possemato & Andrea Gazzarini

https://share-family.org
https://wiki.svde.org/
https://.svde.org
info@svde.org

62. Congresso Nazionale AIB
Biblioteche e tecnologie al tempo dell'Intelligenza Artificiale
Firenze - 16-17 novembre 2023