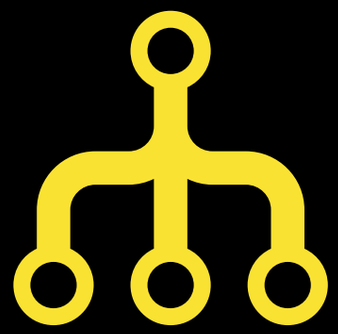
WODHSA2019

The Literary Theme Ontology for Media Annotation and Information Retrieval

Paul Sheridan, Tupac.Bio, Japan/USA
Mikael Onsjö, Independent Researcher, UK
Janna Hastings, University College London, UK

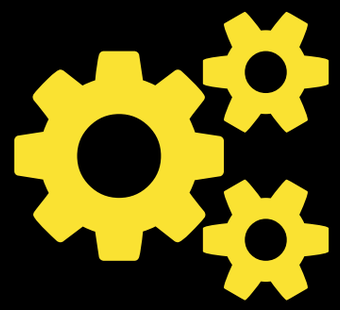
What is the LTO?



An ontology of literary themes

An online collaborative fiction document thematic annotation platform

An open access database of thematically annotated stories



An ecosystem of statistical analysis applications

Where is the LTO?

theme ontology github

theme ontology homepage

Search or jump to... / Pull requests Issues Marketplace Explore

theme-ontology / theming

Watch 1 Star 0 Fork 0

Code Issues 0 Pull requests 0 Projects 0 Wiki Security Insights Settings

The core repository for the Theme Ontology project. Edit

Manage topics

3,378 commits 2 branches 7 releases 2 contributors CC-BY-4.0

Branch: master New pull request Create new file Upload files Find File Clone or download

paul-sheridan The Disappearance of Flight 412 (1974) initial review Latest commit b89818e 1 hour ago

docs	added a few German language films	2 hours ago
notes	The Disappearance of Flight 412 (1974) initial review	1 hour ago
LICENSE.md	CC license initial import	22 days ago
README.md	Update README.md	8 days ago

README.md

Theming

https://www.themeontology.org/

What is a Story?

Any publicly known narrative that can be referenced and analyzed may be a "story" in this project. To be admissible, a story entry must merely:

- Unambiguously identify a story from the data entered, chiefly by date and title.
- Indicate a story that's obtainable to anyone who wishes to review the themes of it.
- Come with a reasonable expectation that the previous conditions will continue to hold for a very long time.

Conventional published sources such as books and movies are nearly always acceptable. Any determined themer can purchase their copy from one vendor or another. Some examples of works that are not admissible have emerged: Lost works (commentary may exist that ellucidate themes but we have no way of reviewing the work for ourselves) are not accepted. Works, such as many school essays, which may only be obtainable via some url online are generally not acceptable since urls frequently become broken.

What is a Theme?

A theme is any topic of interest that is featured in a story. A central theme is a theme that is topical throughout most of the story, either because it is featured continously or because it pertains to the main storyline in its entirity. For example a "moral of the story" may be revealed only towards the end, yet is considered central because it is what most of the story was leading up to. A peripheral theme is any theme of interest that is featured briefly

THEMEONTOLOGY.ORG Stories Themes Resources Login

Fuzzy Search:

ID	Title	Date	Description
Collection: All	All Stories	various	All defined stories in the database.
Collection: Miscellaneous	Miscellaneous	various	Otherwise uncategorized stories in the database.
Collection: Star Trek: Discovery	Star Trek: Discovery	2017–	Star Trek: Discovery is an American television series created for CBS All Access by Bryan Fuller and Alex Kurtzman. It is the first series developed specifically for that service, and the first Sta...
Collection: Electric Dreams	Electric Dreams	2017	Philip K. Dick's Electric Dreams, or simply Electric Dreams, is a science fiction television anthology series based on the works of Philip K. Dick. The series consists of ten standalone episodes ba...
Collection: Black Mirror	Black Mirror	2011–	Black Mirror is a British anthology science fiction television series created by Charlie Brooker, with Brooker and Annabel Jones serving as the programme showrunners. It examines modern society, pa...
Collection: Game of Thrones	Game of Thrones	2011	Game of Thrones is an American fantasy drama television series created by David Benioff and D. B. Weiss. It is an adaptation of A Song of Ice and Fire, George R. R. Martin's series of fantasy novel...
Collection: Science fiction films of the 2010s	Science fiction films of the 2010s	2010–2019	Science fiction films that premiered between 1 January 2010 and 31 December 2019. References: https://en.wikipedia.org/wiki/List_of_science_fiction_films_of_the_2010s
Collection: Sherlock	Sherlock	2010–2017	Sherlock is a British crime drama television series based on Sir Arthur Conan Doyle's Sherlock Holmes detective stories. Created by Steven Moffat and Mark Gatiss, it stars Benedict Cumberbatch as S...
Collection: Star Trek: Enterprise	Star Trek: Enterprise	2001–2005	Star Trek: Enterprise (ENT) (titled simply Enterprise until the third episode of season three) is an American science fiction television series created by Rick Berman and Brannon Braga as a prequel...
Collection: X-Men	X-Men	2000–2017	X-Men is an American superhero film series based on the fictional superhero team of the same name, who originally appeared in a series of comic books created by Stan Lee and Jack Kirby and publishe...

Show 10 entries filter:

Showing 1 to 10 of 51 entries Previous 1 2 3 4 5 6 Next

stories.php workflow | util | log | uat | donate

2008-2016

2017

2018

2019

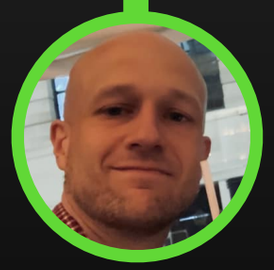
Paul

GitHub

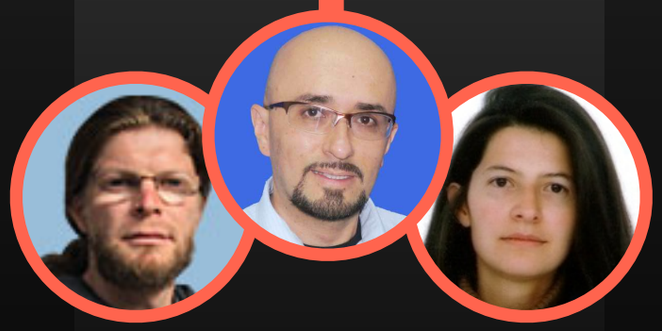
WODHSA

DSCN

Janna

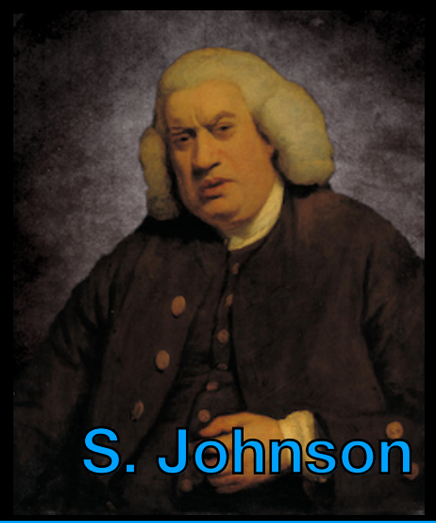


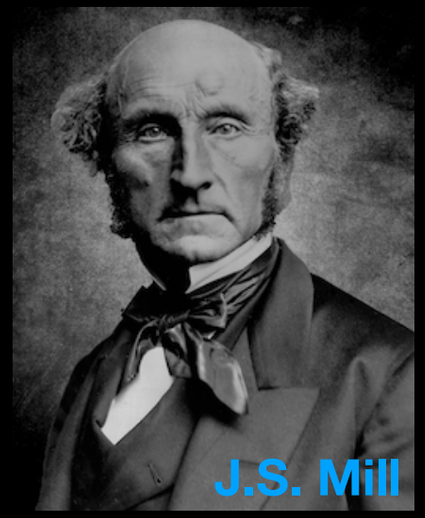
Mikael

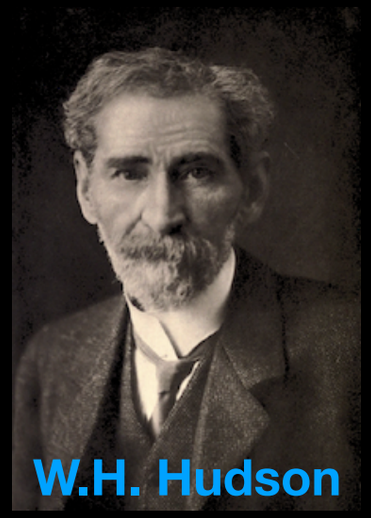


Sergio Team

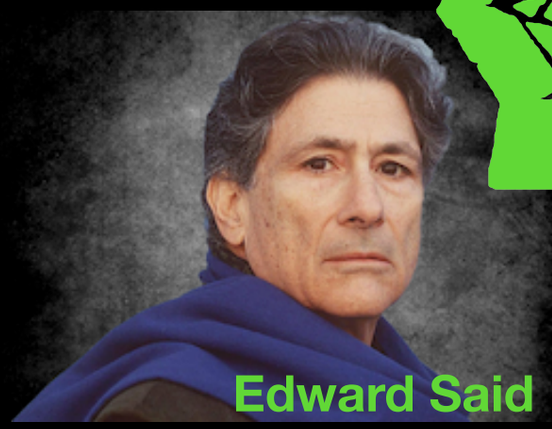
Future Internet







The Gentlemen Themers



How is the LTO Structured?

William Henry Hudson
(1841-1922)

literary thematic entity

the human condition

- **human individual**
- **human pair**
- **human group**
- **human society**

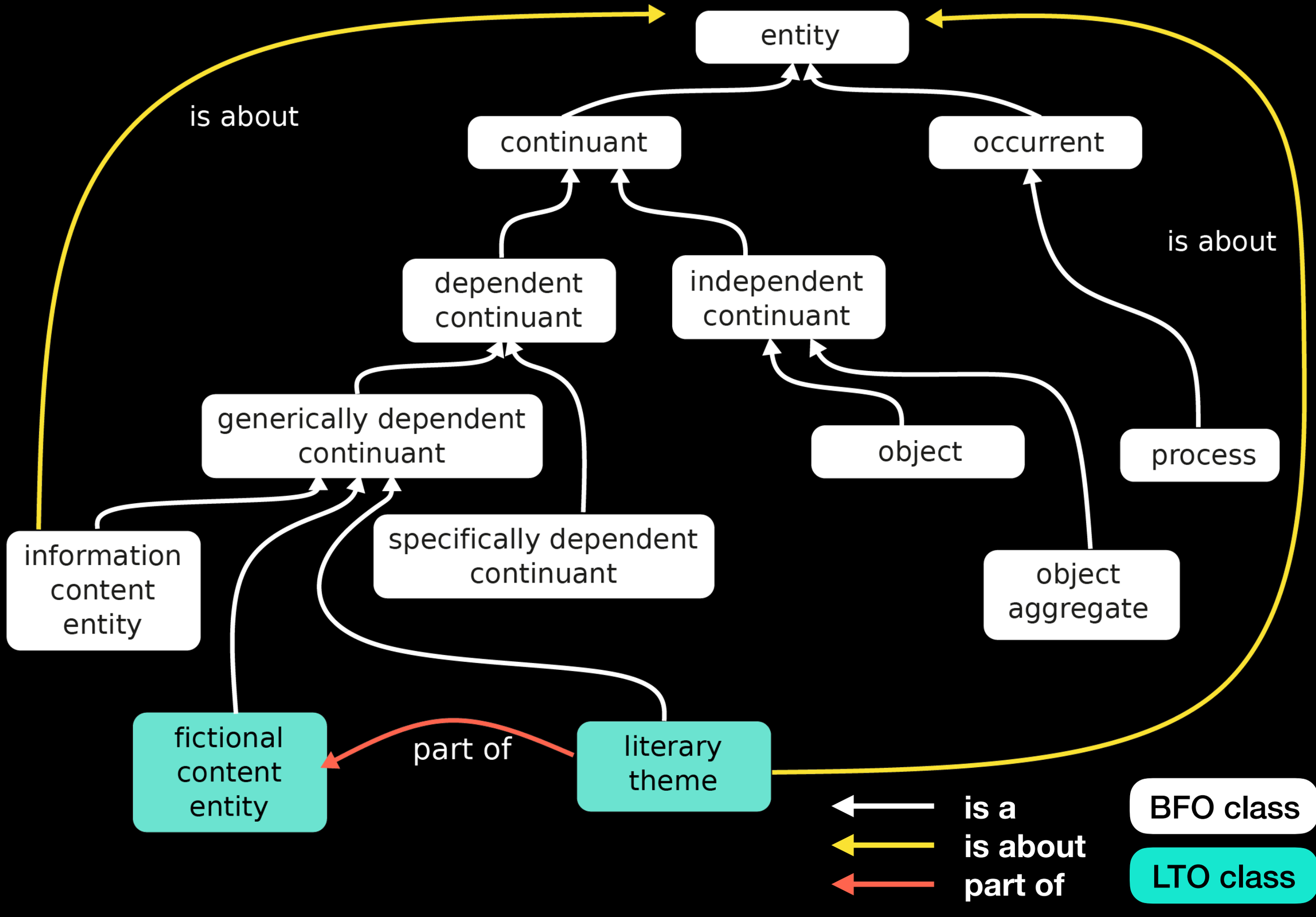
the pursuit of knowledge

- **the art of war**
- **science and philosophy**
- **religion and the supernatural**

speculative fiction theme

- **speculative being**
- **extraordinary experience**
- **speculative technology**
- **speculative society**

Alignment with Basic Formal Ontology



Example: Vengeance in Hamlet

play: Hamlet (1602) story

Title: Hamlet
Date: 1602

The Tragedy of Hamlet, Prince of Denmark, often shortened to Hamlet (/ˈhæmlɪt/), is a tragedy written by William Shakespeare sometime between 1599 and 1602. Set in Denmark, the play depicts Prince Hamlet and his revenge against his uncle, Claudius, who has murdered Hamlet's father in order to seize his throne and marry Hamlet's mother.

References:
https://en.wikipedia.org/wiki/Hamlet

the desire for vengeance theme

A character seeks retribution over a perceived injury or wrong.

References:
https://en.wikipedia.org/wiki/Revenge

literary thematic entity

the human condition

human individual

human need

human self-esteem need

the desire for vengeance

Theme	Weight	Motivation
the desire for vengeance	choice	Hamlet sought revenge against his uncle, Claudius, who had murdered Hamlet's father in order to seize his throne and marry Hamlet's mother Gertrude.

The HSRB Tagging Protocol | Identify an important topic in the story: who or what is at the center of that topic?

humanity

- a single individual
  - their emotional states → anger, boredom, fear, grief, hope love, shame, etc.
  - their challenges in life → coping with death, tough decision, coping with aging, etc.
  - their purpose in life → the lust for gold, the desire for vengeance, the quest for immortality, the thirst for knowledge, etc.
  - introspection → seeing the error of one's ways, having a preconception challenged, overcoming a limitation, etc.
- a pair of individuals → a familial relationship, a friendship, a romantic relationship, etc.
- a small group of individuals → family life, a workplace issue, a social status, etc.

society

- a society's internal matter → a gender issue, a socioeconomic issue, a law and order issue, etc.
- an inter-societal matter → diplomacy, war, cross cultural issue, etc.
- a universal societal matter → human rights, social ethical dilemma, social organization, etc.

reality

- knowledge of reality through faith → religion, the supernatural, etc.
- knowledge of reality through reason → philosophy, science, the arts, ethics, etc.

...and beyond

- a fantastical individual experience → what if I found myself in a different time/place, what if I had an extraordinary ability, etc.
- a speculative society → utopia, dystopia, existential risk to civilization, etc.
- speculative technologies → AI, human enhancements, speculative devices, simulated reality, etc.
- crackpot theories → alien abduction, Atlantis, Hollow Earth, time traveler secretly living among us, etc.

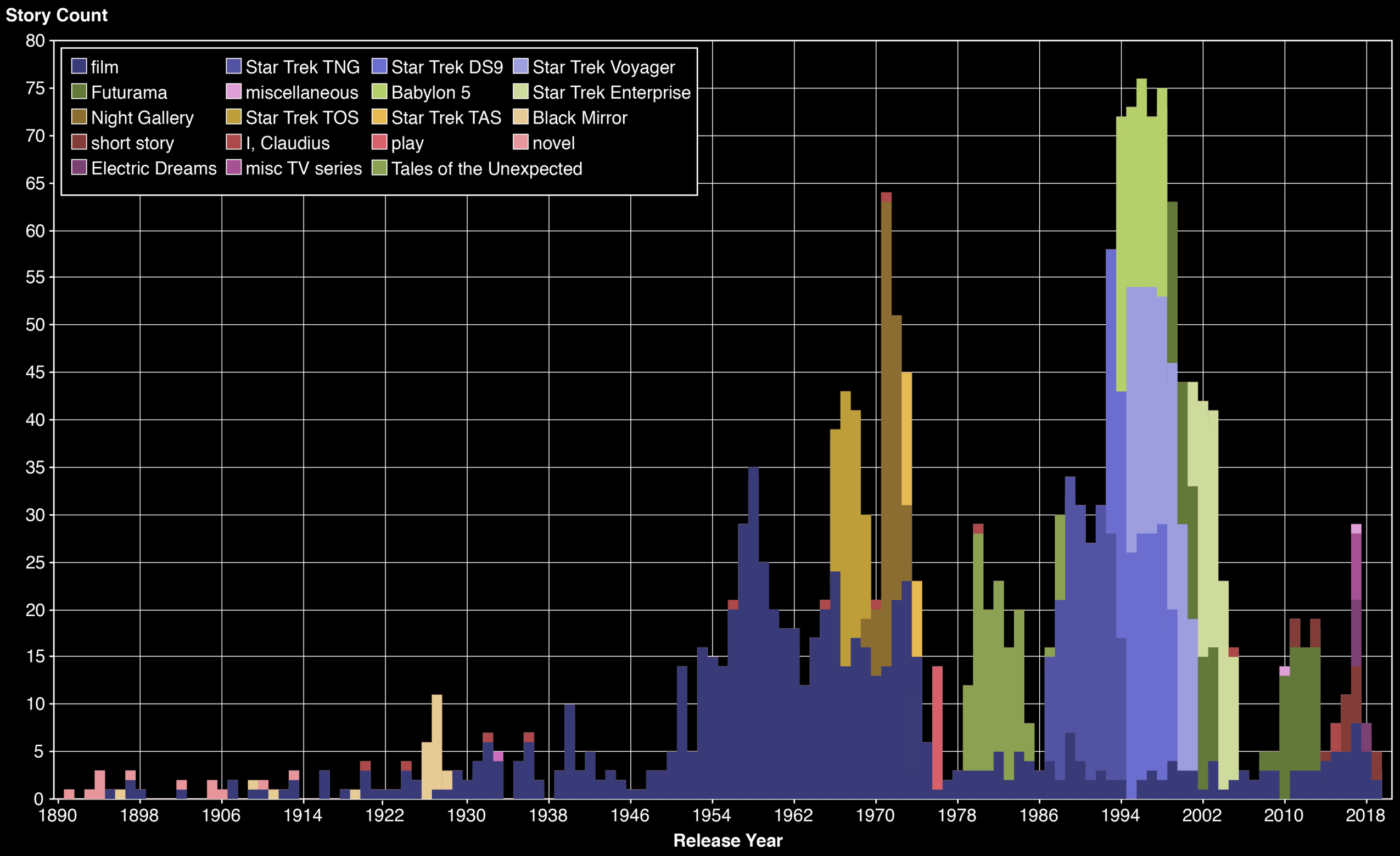
Look up specific themes at https://www.themeontology.org/themes.

LTO Theme Acquisition

Story Collection	Time Period	Story Count
Classic sci-fi films	1895-1974	500
Modern sci-fi films	1975-2019	100
Star Trek: The Original Series	1966-1969	79
Star Trek: The Animated Series	1973-1974	22
Star Trek: The Next Generation	1987-1994	178
Star Trek: Deep Space Nine	1993-1999	177
Babylon 5	1994-1998	110
Star Trek: Voyager	1995-2001	172
Futurama	1999-2013	129
Star Trek: Enterprise	2001-2005	99
Black Mirror	2011-2019	23
Other	1602-2019	400

Total = 2000

LTO Theme Acquisition



Application: Story Recommender

future internet

MDPI

Article

An Ontology-Based Recommender System with an Application to the Star Trek Television Franchise

Paul Sheridan [1,*], Mikael Onsjö [2], Claudia Becerra [3], Sergio Jimenez [4] and George Dueñas [4]

[1] Tupac Bio, Inc., San Francisco, CA 94103, USA
[2] Independent Researcher, London, SE13 7NZ, UK
[3] Systems and Computer Engineering Department, Universidad Nacional de Colombia, Ciudad Universitaria bldg. 453, Bogotá, D.C. 11001, Colombia
[4] Instituto Caro y Cuervo, Calle 10 # 4-69, Bogotá, D.C. 111711, Colombia
* Correspondence: paul.sheridan.stats@gmail.com

Received: 8 July 2019; Accepted: 14 August 2019; Published: 22 August 2019

Abstract: Collaborative filtering based recommender systems have proven to
in settings where user preference data on items is abundant. However
algorithms are hindered by their weakness against the item cold-start prob
interpretability. Ontology-based recommender systems exploit hierarchica
and items to enhance browsing, recommendation, and profile construction
approaches address the shortcomings of their collaborative filtering co
organizations of items can be difficult to obtain for items that mostly belo
(e.g., television series episodes). In this paper, we present an ontology-base
that integrates the knowledge represented in a large ontology of literary th
content recommendations. The main novelty of this work is an ontology-base
similarities between items and its integration with the classical Item-KN
algorithm. As a study case, we evaluated the proposed method again
performing the classical rating prediction task on a collection of Star Trek tele
an item cold-start scenario. This transverse evaluation provides insights in
information resources and methods for the initial stages of recommende
We found our proposed method to be a convenient alternative to collaborat
for collections of mostly similar items, particularly when other content-ba
applicable or otherwise unavailable. Aside from the new methods, this pap
for future research and an online framework to collaboratively extend the on
to cover other narrative content.

Keywords: knowledge-based recommender systems; knowledge represe
ontological engineering; ontology population; ontology-based recommende

1. Introduction

Recommender systems (RSs), or recommenders for short, help users to navigate large collections of items in a personalized way [1]. Broadly speaking, RSs function by correlating user preference data with item ... to generate ... list of rec... items for ... Systems of ... to the ... gs to ... like'

Story recommender

Choose a story:
voy3x05: False Profits

Theme levels:
- Central and Peripheral
- Central
- Peripheral

Similarity function:
cosine

Minimum theme overlap:
1

Background storyset (SMT file format):
Browse... StarTrek.smt
Upload complete

Blacklisted themes (input one theme per line):

Reset

Show 10 entries Search:

Story ID	Title	Score	Common Themes	Count
tng4x13	Devil's Due	0.588	avarice, exploitation of sentient beings, fraud, religion as a control mechanism, the fulfillment of prophesy, the lust for gold	6
tos2x20	A Piece of the Action	0.401	avarice, the ethics of interfering in less advanced societies, the lust for gold	3
tng1x22	Symbiosis	0.343	avarice, exploitation of sentient beings, the ethics of interfering in less advanced societies, the lust for gold	4
tos2x25	The Omega Glory	0.320	avarice, the ethics of interfering in less advanced societies, the lust for gold	3
tas1x10	Mudd's Passion	0.289	avarice, the lust for gold	2
tng6x18	Starship Mine	0.267	avarice, the lust for gold	2
tng3x04	Who Watches The Watchers	0.250	primitive point of view, the ethics of interfering in less advanced societies	2
voy7x06	Inside Man	0.250	avarice, the lust for gold	2
tng5x09	A Matter of Time	0.236	avarice, the lust for gold	2
voy7x14	Prophecy	0.236	religion as a control mechanism, the fulfillment of prophesy	2

Showing 1 to 10 of 44 entries Previous 1 2 3 4 5 Next

Theme Enrichment Analysis: A Statistical Test for Identifying Significantly Enriched Themes in a List of Stories with an Application to the Star Trek Television Franchise

Mikael Onsjö [1] **and Paul Sheridan** [2,*]

[1] *Independent Researcher*
[2] *Department of Active Life Promotion Science, Graduate School of Medicine, Hirosaki University, 5 Zaifu, Hirosaki, Aomori, Japan, 036-8562*

Correspondence*:
Paul Sheridan
paul.sheridan.stats@gmail.com

Digital Studies / Le champ numérique

ABSTRACT

In this paper, we describe how the hypergeometric test can be used to determine whether a given theme of interest occurs in a storyset at a frequency more than would be expected by chance. By a storyset we mean simply a list of stories defined according to a common attribute (e.g., author, movement, period). The test works roughly as follows: Given a background storyset and a sub-storyset of interest, the test determines whether a given theme is over-represented in the sub-storyset, based on comparing the proportions of stories in the sub-storyset and background storyset featuring the theme. A storyset is said to be "enriched" for a theme with respect to a particular background storyset, when the theme is identified as being significantly over-represented by the test. Furthermore, we introduce here a toy dataset consisting of 280 manually themed Star Trek television franchise episodes. As a proof of concept, we use the hypergeometric test to analyze the Star Trek stories for enriched themes. The hypergeometric testing approach to theme enrichment analysis is implemented for the Star Trek thematic dataset in the R package **stoRy**. A related R **Shiny** web application can be found at `https://github.com/theme-ontology/shiny-apps`.

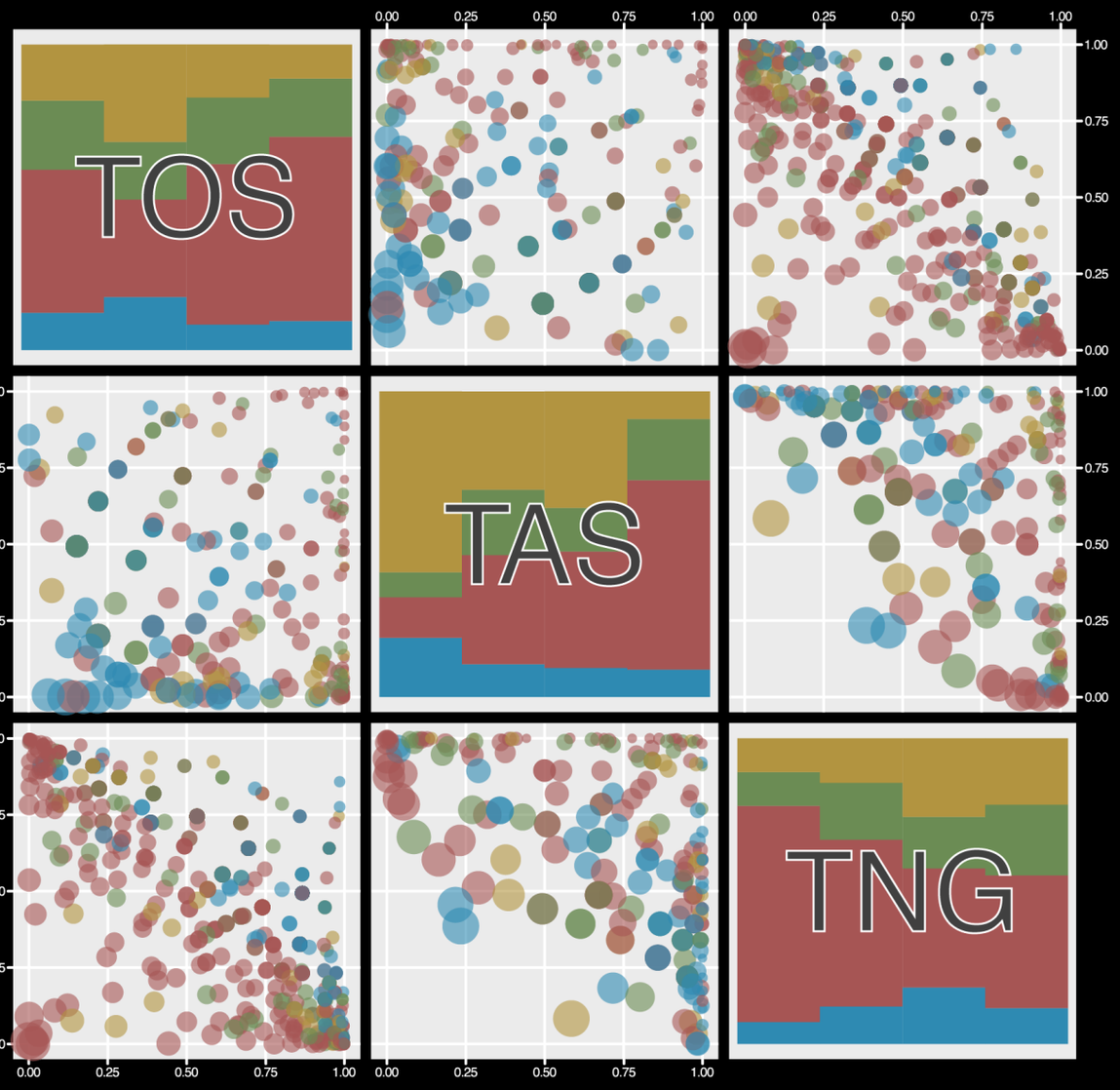
Keywords: enrichment analysis, hypergeometric test, over-representation analysis, Star Trek, theme ontology

1 INTRODUCTION

A *literary theme*, or *theme* for short, is loosely defined as "An idea that recurs in or pervades a work of art or literature" (e.g., a story) [1]. Themes are often expressible in a single word or short phrase, as is illustrated by such garden-variety themes as "love", "loyalty", and "the lust for gold". These examples all happen to be value-neutral abstractions. But themes can just as well take the form of morally charged messages, such as "be wary of strangers" and "do not judge prematurely". The consummate story-maker usually takes pains to imply a theme indirectly, rather than state it explicitly. Sometimes the story-maker is even unconscious of important themes found in their stories. A typical story will feature multiple themes. In the [illegible] *central* themes (i.e., themes found to recur throughout a major part [illegible] conclusion) and *peripheral* themes (i.e., briefly featured themes that

[1] https://en.oxforddictionaries.com/definition/theme/ [Online; accessed

Application: Theme Enrichment Analysis



Future Work

- Publish fiction document clustering method.
- Publish time-course analysis method.
- Publish method to sync topic model topics with LTO themes.
- Collaborate with literary scholars to improve LTO structure and themes.
- Publish fiction document LTO thematic annotation protocol.
- Build up online community.
- Build up annotated fiction document database.

Thank You