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?

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 meet:

- Unambiguously identify a story from the date 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 elucidate 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 continuously or because it pertains to the main storyline in its entirety. 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 is leading up to. A similar problem is any theme of interest that is featured briefly.
How is the LTO Structured?

- 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

William Henry Hudson (1841-1922)
Alignment with Basic Formal Ontology

entity

continuant

dependent continuant

generically dependent continuant

information content entity

specifically dependent continuant

literary theme

occurred

independent continuant

object

object aggregate

process

fictional content entity

is about

is a

part of

BFO class

LTO class
Example: Vengeance in Hamlet

The Tragedy of Hamlet, Prince of Denmark, often shortened to Hamlet (ˈhæmlɛt), 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

A character seeks retribution over a perceived injury or wrong.

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

<table>
<thead>
<tr>
<th>Theme</th>
<th>Weight</th>
<th>Motivation</th>
</tr>
</thead>
<tbody>
<tr>
<td>the desire for vengeance</td>
<td>choice</td>
<td>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.</td>
</tr>
</tbody>
</table>
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.
- 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.

...and beyond

Look up specific themes at https://www.themeontology.org/themes.
## LTO Theme Acquisition

<table>
<thead>
<tr>
<th>Story Collection</th>
<th>Time Period</th>
<th>Story Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classic sci-fi films</td>
<td>1895-1974</td>
<td>500</td>
</tr>
<tr>
<td>Modern sci-fi films</td>
<td>1975-2019</td>
<td>100</td>
</tr>
<tr>
<td>Star Trek: The Original Series</td>
<td>1966-1969</td>
<td>79</td>
</tr>
<tr>
<td>Star Trek: Deep Space Nine</td>
<td>1993-1999</td>
<td>177</td>
</tr>
<tr>
<td>Babylon 5</td>
<td>1994-1998</td>
<td>110</td>
</tr>
<tr>
<td>Star Trek: Voyager</td>
<td>1995-2001</td>
<td>172</td>
</tr>
<tr>
<td>Futurama</td>
<td>1999-2013</td>
<td>129</td>
</tr>
<tr>
<td>Star Trek: Enterprise</td>
<td>2001-2005</td>
<td>99</td>
</tr>
<tr>
<td>Black Mirror</td>
<td>2011-2019</td>
<td>23</td>
</tr>
<tr>
<td>Other</td>
<td>1602-2019</td>
<td>400</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>2000</strong></td>
<td></td>
</tr>
</tbody>
</table>
Application: Story Recommender
Application: Theme Enrichment Analysis

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 Sheridan2*
1 Independent Researcher
2 Department of Active Life Promotion Science, Graduate School of Medicine, Hiroaki University, 5 Zaitu, Hiroaki, 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 story set at a frequency more than would be expected by chance. By a story set 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 story set and a sub-story set of interest, the test determines whether a given theme is over-represented in the sub-story set, based on comparing the proportions of stories in the sub-story set and background story set featuring the theme. A story set is said to be “enriched” for a theme with respect to a particular background story set, 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 shiny-theme. 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). Themes are often expressible in a single word or short phrase, as is illustrated by such garden variety themes as “love”, “loyalty”, and “the last 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-teller usually takes pains to imply a theme indirectly, rather than state it explicitly. Sometimes the story-teller is even unconcerned about themes found in their stories. A typical story will feature multiple themes. In the best stories, these themes are central themes (i.e., themes found to recur throughout a major part of the story) and peripheral themes (i.e., briefly featured themes that

*https://en.wikipedia.org/wiki/Theme (Online: accessed...
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