



Граф знаний: мифология Говарда Лавкрафта

Станислав Елькин,
магистрант 2 курса «Искусственный интеллект»

ИТМО 2024

Цель и потенциальные пользователи

❖ **Цель:** создание многофункционального графа для глубокого анализа и визуализации мифологии Лавкрафта.

❖ **Потенциальные пользователи:**

- ❖ **Литературоведы:** анализ влияния Лавкрафта на современную литературу.
- ❖ **Создатели контента:** писатели, разработчики игр, сценаристы.
- ❖ **Фанаты:** детальное исследование миров Лавкрафта.
- ❖ **Преподаватели и студенты:** использование для исследовательских и учебных целей.

Основные функции

★ Единая структурированная база данных:

- Все персонажи, распределенные по классам и подклассам.
- Реальные люди, связанные с Лавкрафтом или вдохновленные его творчеством.
- Список локаций, организаций (например, культов) и артефактов, задействованных в произведениях.
- Огромная база произведений, фильмов, игр и других медиа, на которые оказал влияние Лавкрафт.

Компетентностные вопросы

1. Ознакомление с Великими Древними и персонажами мифологии Лавкрафта:

- **Вопрос:** Каков полный список Великих Древних и других персонажей мифологии Лавкрафта? В каких произведениях впервые упоминались Великие Древние?
- **Задача графа:** Позволить пользователю получить список Великих Древних, исследовать их появление в произведениях и изучить структуру мифологии.
- **Пример юзкейса:** Исследователи могут использовать граф для изучения связи персонажей с литературными источниками.

Компетентностные вопросы

2. Связь артефактов с персонажами и произведениями

- **Вопрос:** В каких произведениях и кем использовались артефакты, описанные в мифологии Лавкрафта?
- **Задача графа:** Установить связь между артефактами, персонажами, которые их использовали, и произведениями, где артефакты впервые появились. Это поможет исследовать роль артефактов в сюжете и их значение в контексте мифологии.
- **Пример юзкейса:** Геймдизайнеры могут использовать эти данные для создания карт или игровых миров, связанных с мифологией.

Компетентностные вопросы

3. Персонажи и организации

- **Вопрос:** Какие персонажи связаны с различными культурами, организациями или группами, которые напрямую или косвенно соприкасались с мифологией Лавкрафта? Сколько членов имеет каждая из этих организаций?
- **Задача графа:** Создать список организаций и участников, включая как непосредственно связанные с мифологией Лавкрафта, так и окромиофологические группы.
- **Пример юзкайса:** Литературные исследователи могут анализировать, как мифология Лавкрафта повлияла на другие вымышленные вселенные.

Компетентностные вопросы

4. Влияние мифологии Лавкрафта на международную аудиторию.

- **Вопрос:** На каких языках, кроме английского, были написаны произведения, связанные с мифологией Лавкрафта? Сколько произведений существует на каждом из этих языков? Какой язык является лидером по количеству произведений в текущей онтологии?
- **Задача графа:** Выявить и классифицировать произведения, написанные на других языках, определить лидирующий язык по количеству произведений для анализа влияния мифологии Лавкрафта на международную аудиторию.
- **Пример юзкейса:** Издатели и переводчики могут использовать эти данные для приоритизации переводов на наиболее популярные языки, а культурные аналитики – для оценки глобального культурного влияния Лавкрафта.

Источники информации

★ Fandom (<https://lovecraft.fandom.com/>) – содержит более 5000 существ, стал основным источником для графа знаний.

Аспект	Описание
Структура данных	Четкая структура страниц, включая инфобоксы, использованные для свойств графа.
Подклассы	Хэдеры в начале страниц помогли определить подклассы существ.
Обогащение	Каждая сущность сопровождается аннотацией, что позволило автоматически обогатить базу данных.
Преимущества	Богатая структура и широкое покрытие мифологии.
Недостатки	Наличие мусорных данных, неуниверсальная структура, ограниченная информация о личности Лавкрафта, отсутствие выделенного подкласса существ Great Ones.

★ H.P. Lovecraft's official site (<https://www.hplovecraft.com/life/>):

Этот раздел был использован для обогащения данных о друзьях и родственниках Лавкрафта. Дополняет информацию из Fandom более личными деталями, которые были недоступны на основном ресурсе.

Логика выделения классов и подклассов

❖ Анализируя структуру данных на ресурсе Fandom, было принято решение выделить 6 основных классов для построения онтологии:

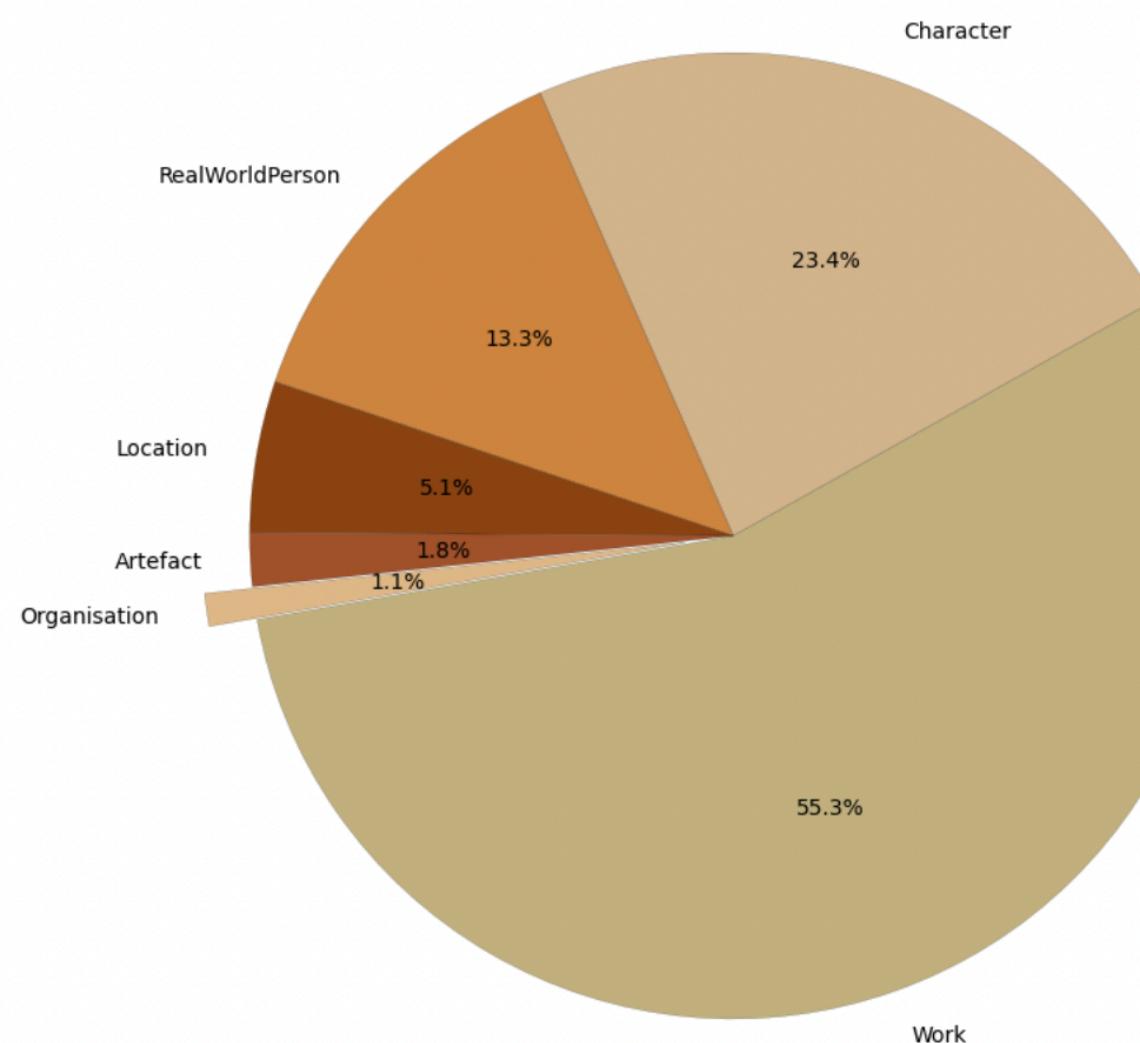
- **Work** (литературные произведения, игры, фильмы);
- **Character** (мифологические существа, герои);
- **RealWorldPerson** (реальные люди);
- **Location** (места, как реальные, так и вымышленные);
- **Artefact** (артефакты, связанные с мифологией);
- **Organisation** (группы, такие как культуры и организации).

Всего обработано **5560 сущностей**. Для каждого класса были автоматически определены подклассы, используя хэдеры страниц.

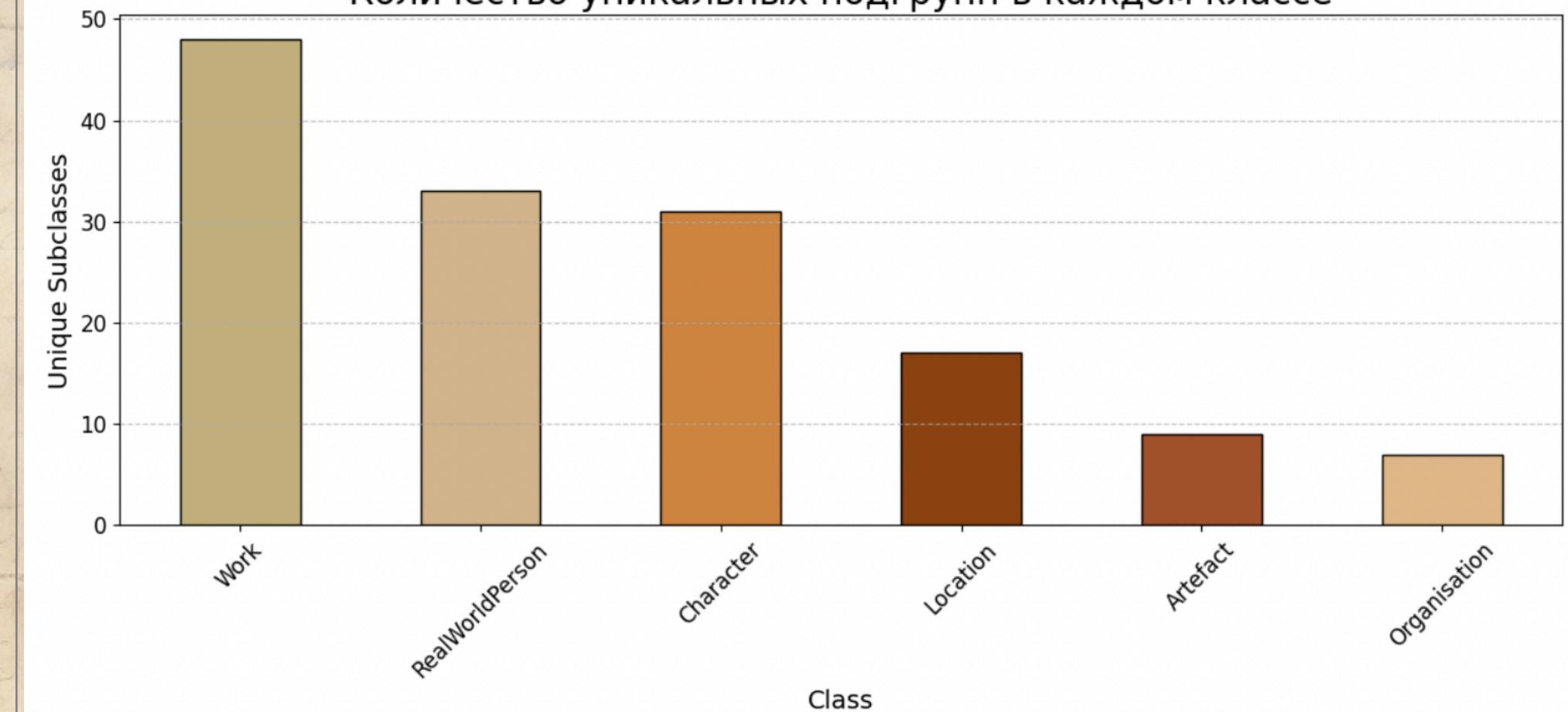
Например, для класса **Character** выделены подклассы, такие как **Elder Gods**, **Great Old Ones**, **Humans** и т.д. Такой подход позволил структурировать данные и сделать онтологию более детализированной.

Логика выделения классов и подклассов

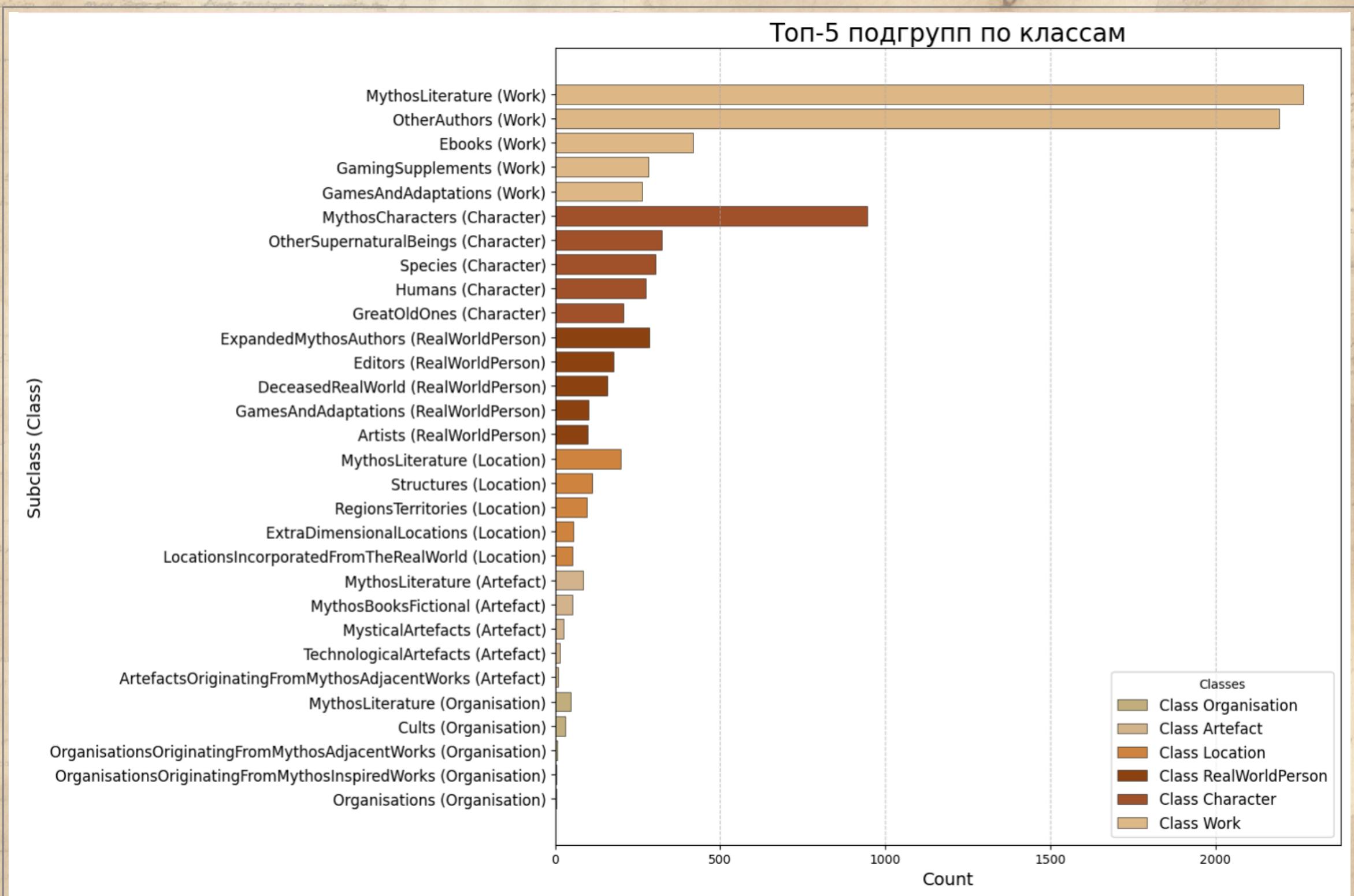
Распределение классов



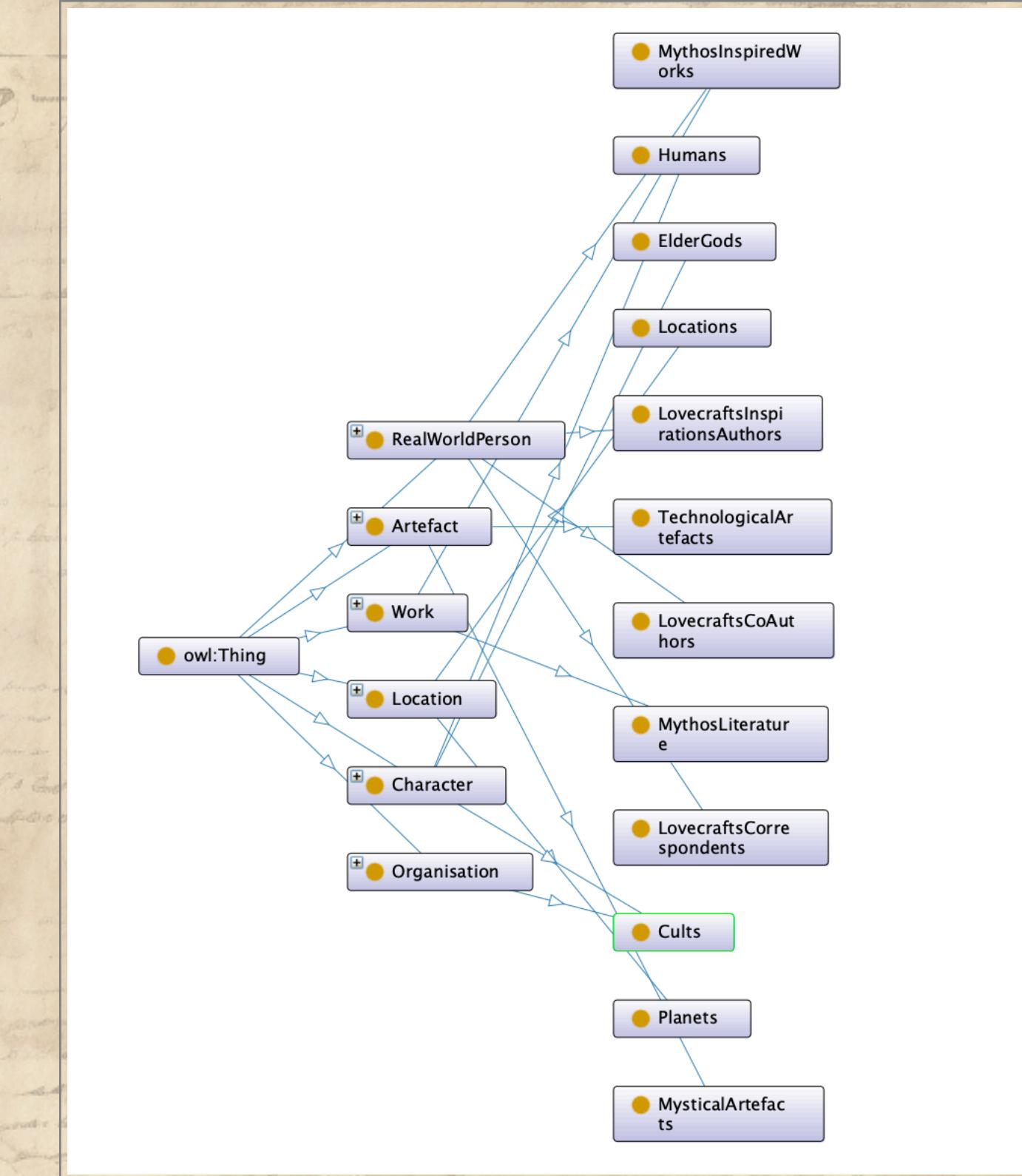
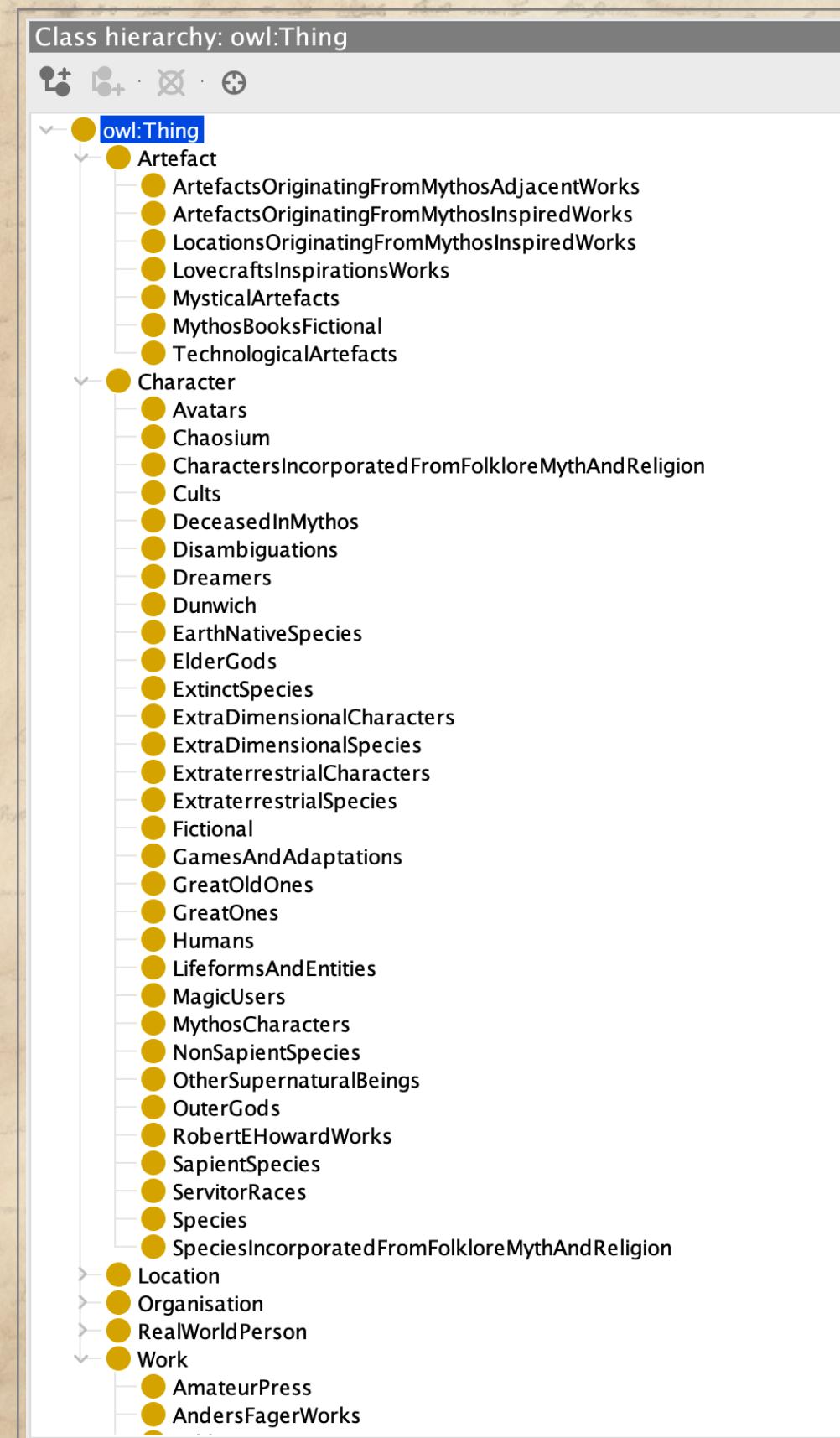
Количество уникальных подгрупп в каждом классе



Логика выделения классов и подклассов



Представление иерархии в графе



О граfe

Ontology metrics:	
Metrics	
Axiom	54 353
Logical axiom count	35 769
Declaration axioms count	248
Class count	136
Object property count	80
Data property count	35
Individual count	5 560
Annotation Property count	43
Class axioms	
SubClassOf	148
EquivalentClasses	0
DisjointClasses	0
GCI count	0
Hidden GCI Count	0
Object property axioms	
SubObjectPropertyOf	0
EquivalentObjectProperties	0
InverseObjectProperties	27
DisjointObjectProperties	0
FunctionalObjectProperty	11
InverseFunctionalObjectProperty	0
TransitiveObjectProperty	6
SymmetricObjectProperty	9
AsymmetricObjectProperty	0
ReflexiveObjectProperty	0
IrreflexiveObjectProperty	0
ObjectPropertyDomain	80
ObjectPropertyRange	111
SubPropertyChainOf	0
Data property axioms	
SubDataPropertyOf	0
EquivalentDataProperties	0
DisjointDataProperties	0
FunctionalDataProperty	13
DataPropertyDomain	32
DataPropertyRange	35

Individual axioms

ClassAssertion	18 077
ObjectPropertyAssertion	5 238
DataPropertyAssertion	11 982
NegativeObjectPropertyAssertion	0
NegativeDataPropertyAssertion	0
SameIndividual	0
DifferentIndividuals	0

Annotation axioms

AnnotationAssertion	18 336
AnnotationPropertyDomain	0
AnnotationPropertyRangeOf	0

Определение свойств (properties)

❖ На основе анализа инфобоксов страниц и их содержания были определены основные свойства для каждого класса. Например:

- **Character**: hasOrigin, hasRelatives, hasAppearance.
- **Work**: hasAuthor, hasGenre, wasPublishedOn.
- **Artefact**: isLocatedAt, hasType, madeFirstAppearanceOn.
- **RealWorldPerson**: hasBirthplace, hasOccupation, hasRelatives.
- **Location**: isLocatedAt, hasAccessibility, hasNotableFeature.
- **Organisation**: hasAffiliation, hasPurpose, hasNotableMember.

Часть свойств была исключена, чтобы избежать избыточности и сделать онтологию более понятной. При дальнейшем развитии проекта возможно добавление новых свойств и уточнение существующих.

Обоснование классификации свойств

Data Properties и Object Properties:

❖ **Data Properties** были назначены для свойств, которые связывают сущности с их атрибутами, представленными в виде простых значений (строк, чисел, дат). Например:

- hasGenre для **Work** — это строковое значение, описывающее жанр произведения.
- wasBornOn для **RealWorldPerson** — это дата рождения, хранимая в формате строки.

Причина выбора: Data Properties используются для представления атомарных данных, которые не требуют связи с другими сущностями.

❖ **Object Properties** были назначены для свойств, описывающих связи между сущностями графа. Например:

- hasAuthor для **Work**, связывающее произведение с автором (другой сущностью **RealWorldPerson**).
- isLocatedAt для **Location**, связывающее локацию с другой сущностью (например, артефактом или персонажем).

Причина выбора: Object Properties используются для отображения отношений между разными классами сущностей.

Object property hierarchy: hasRealPersonMateR

Asserted

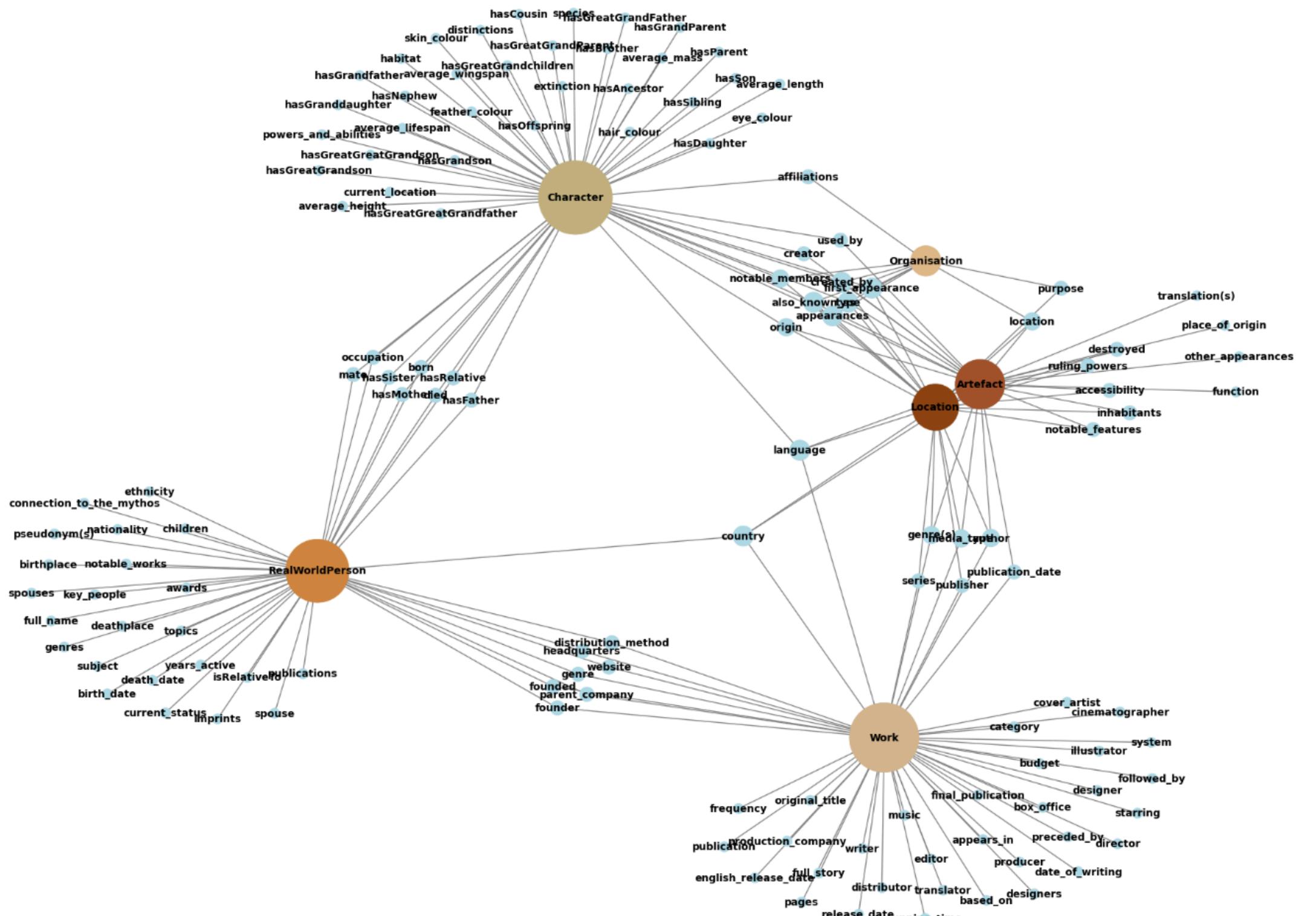
- owl:topObjectProperty
 - created
 - hasAffiliationC
 - hasAffiliationO
 - hasAffiliationR
 - hasAffiliationW
 - hasAncestor
 - hasAncestorR
 - hasAppearanceA
 - hasAppearanceC
 - hasAppearanceL
 - hasAppearanceW
 - hasAuntR
 - hasAuthorW
 - hasBirthplaceR
 - hasBrotherR
 - hasCharacterMateC
 - hasChild
 - hasChildR
 - hasChildrenR
 - hasCountryOfOriginW
 - hasCousinR
 - hasDaughterR
 - hasDescendant
 - hasDescendantR
 - hasFatherR
 - hasGrandChild
 - hasGrandChildR
 - hasGrandDaughterR
 - hasGrandFatherR
 - hasGrandMotherR
 - hasGrandParent
 - hasGrandSonR
 - hasHalfBrotherR
 - hasHalfSibling
 - hasHalfSisterR
 - hasMotherR
 - hasNationalityR
 - hasNephewOrNiece
 - hasNephewR
 - hasNieceR
 - hasNotableMemberO

Data property hierarchy:

Asserted

- owl:topDataProperty
 - ceasedOnC
 - diedOnR
 - hasAccessibilityL
 - hasAlternativeNamesA
 - hasAlternativeNamesC
 - hasAlternativeNamesL
 - hasAlternativeNamesO
 - hasAlternativeNamesR
 - hasDistinctionC
 - hasFullNameR
 - hasFunctionA
 - hasGenreW
 - hasNotableFeatureL
 - hasOccupationR
 - hasOriginA
 - hasOriginL
 - hasPageCountW
 - hasPlaceOfOriginA
 - hasPseudonymR
 - hasPurposeL
 - hasPurposeO
 - hasTypeA
 - hasTypeL
 - hasTypeO
 - hasWebsiteR
 - hasYearsActiveR
 - isInLanguageW
 - isPublishedAsW
 - usesLanguageC
 - wasBornOnR
 - wasDestroyedOnA
 - wasDestroyedOnL
 - wasPublishedOnW
 - wasReleasedOnW
 - wasWrittenOnW

Связи между классами и ключами инфобоксов



Представление индивидуумов

Active ontology x Entities x Individuals by class x DL Query x OntoGraf x SPARQL Query x

Data properties Annotation properties Datatypes Individuals

Classes Object properties

Individuals: Cthulhu

Cth%3A4at_Aquadingen
Cthaat
Cthaat_Aquadingen:_A_Guide_to_Further_Research
Cthaeghya
Cthalpa
Cthugha
Cthulhi
Cthulhu
Cthulhu%3F_Gesundheit%21
Cthulhu:_Death_May_Die
Cthulhu:_Land_of_the_Long_White_Cloud
Cthulhu:_Le_mythe
Cthulhu:_Le_mythe_-_Livre_II
Cthulhu:_Le_mythe_-_Livre_III
Cthulhu:_The_Great_Old_One_%28card_game%29
Cthulhu:_The_Myths_and_Kindred_Horrors
Cthulhu:_Weirdbook_Annual_2
Cthulhu:_Xothic_Wars
Cthulhu_%282007_film%29
Cthulhu_%28Festa%29
Cthulhu_%28Gunship%29
Cthulhu_%28Michael_Fantina%29
Cthulhu_2000:_A_Lovecraftian_Anthology
Cthulhu_4_Kids:_Old_Ones_at_the_Beach:_Volume_1
Cthulhu_4_Kids:_Raising_Rlyeh:_Volume_2
Cthulhu_500
Cthulhu_and_Other_Monsters
Cthulhu_Apocalypse
Cthulhu_Armageddon
Cthulhu_Armageddon_%28series%29
Cthulhu_Britannica
Cthulhu_Britannica:_London
Cthulhu_by_Gaslight
Cthulhu_Casebook
Cthulhu_Christmas_Special_and_Other_Lovecraftian_Yuletide_Tales
Cthulhu_Classics
Cthulhu_Codex_August_1996

Annotations Usage

Annotations: Cthulhu

rdfs:label [in OntologyID(Anonymous-5)]
Cthulhu

hasAffiliationC [in OntologyID(Anonymous-5)]
Great Old Ones Cthulhu Cult Deep Ones

hasAnnotation [in OntologyID(Anonymous-5)]
Also known as see here exp : vhoorl , 23rd nebula outer space current location r'lyeh powers and abilities oneiopathy regeneration shapeshifting immortality relatives hpl : azathoth (great-great-grandfather) the nameless mist (great-grandparent) yog-sothoth (grandfather) shub-niggurath (grandmother) nug (parent)

hasAppearanceC [in OntologyID(Anonymous-5)]
000 Years Under the Sea The Hidden Cthulhu Saves Christmas The Three Stooges vs. Cthulhu Godzilla vs. Cthulhu: A Death May Die Story

Description: Cthulhu

Types +

Character
GreatOldOnes
MythosCharacters

Same Individual As +

Different Individuals +

Property assertions: Cthulhu

Object property assertions +

wasCreatedByC H. P. Lovecraft
isLocatedAtC Rlyeh
madeFirstAppearanceOnC The_Call_of_Cthulhu

Data property assertions +

hasAlternativeNamesC "See here"

Negative object property assertions +

Negative data property assertions +

SPARQL-запросы

- В каких произведениях впервые упоминались Великие Древние?

```

g = Graph()
g.parse("lovecraft_ontology.owl", format="xml")

query_str = """
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ex: <http://example.org/lovecraft#>
SELECT ?character ?work
WHERE {
    ?character rdf:type ex:GreatOldOnes .
    ?character ex:madeFirstAppearanceOnC ?work .
}
"""

results = g.query(query_str)

data = []
for row in results:
    data.append((clean_uri(str(row.character)), clean_uri(str(row.work))))

df = pd.DataFrame(data, columns=["Character", "Work"])

display(df)

```

	Character	Work
0	Sktai	Nightmares_Disciple
1	Animiculi	Beyond_the_Mountains_of_Madness
2	Shuma-Gorath	Appearance - Marvel Premiere #10
3	Shuma-Gorath	Mentioned - The Curse of the Golden Skull
4	Quachil_Uttaus	The Treader of the Dust
...
215	Toth	Usurped
216	Aylith	Twilight Memories
217	Xcthol	Ravenstone Sanitarium (RPG Material)
218	Voltiyig	The Thing in the Library
219	Eilor	Correlated_Contents

220 rows x 2 columns

- В каких произведениях и кем использовались артефакты, описанные в мифологии Лавкрафта?

```

query_str = """
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ex: <http://example.org/lovecraft#>

SELECT ?artefact ?character ?work
WHERE {
    ?artefact rdf:type ex:Artefact .
    ?artefact ex:isUsedByA ?character .
    ?artefact ex:madeFirstAppearanceOnA ?work .
}
"""
|
```

	Artefact	Character	Work
0	Thaumaturgical_Prodigies_in_the_New-England_Ca...	Ambrose Dewart	The_Lurker_at_the_Threshold
1	Vigilant	The Morrison Co.	The_Call_of_Cthulhu
2	Elder_Key_%28Stone_Tablet%29	Ubbo-Sathla	Ubbo-Sathla
3	Ktynga	Fthaggua	The_Fire_Vampires
4	Elder_Key_%28Sigil%29	Haon-Dor	The_Descent_into_the_Abyss
5	Robot_Yeti	The Great Intelligence	The_Abominable_Snowmen
6	Alert	Crew of the Emma	The_Call_of_Cthulhu
7	Alert	Members of the Cthulhu Cult	The_Call_of_Cthulhu
8	Hydrophinnae	Titus_Crow	The_Aquarium
9	Golden_mead	Laban_Shrewsbury	The_House_on_Curwen_Street
10	Golden_mead	Andrew Phelan	The_House_on_Curwen_Street
11	Black_Fan	Nyarlathotep	Masks_of_Nyarlathotep
12	Tulu-metal	K'n-yanians	The_Mound
13	De_Marignys_Clock_%28machine%29	Henri-Laurent_de_Marigny	Through_the_Gates_of_the_Silver_Key
14	De_Marignys_Clock_%28machine%29	Titus_Crow	Through_the_Gates_of_the_Silver_Key
15	Mist_Projector	Mi-Go	Call_of_Cthulhu
16	Crystal_of_the_Elder_Things	Elder_Things	Call_of_Cthulhu_%28Role-Playing_Game%29
17	Clawed_Crystal	Le_Novar	His_Hour_Upon_the_Stage
18	Clawed_Crystal	Robert_Clarke	His_Hour_Upon_the_Stage
19	Book_of_Eibon	Eibon	Ubbo-Sathla
20	Black_Galley	Men_of_Leng	The_Dream-Quest_of_Unknown_Kadath
21	Black_Galley	Moon-beasts	The_Dream-Quest_of_Unknown_Kadath
22	Silver_Key	Randolph_Carter	The_Silver_Key
23	Liao	Halpin_Chalmers	The_Hounds_of_Tindalos
24	Black_Rites	James_Allington	The_Suicide_in_the_Study
25	Mi-Go_Brain_Cylinder	The_Mi-Go	The_Whisperer_in_Darkness

SPARQL-запросы

- Какие персонажи связаны с различными культурами, организациями или группами, которые напрямую или косвенно соприкасались с мифологией Лавкрафта? Сколько членов имеет каждая из этих организаций?

```
query_str = """
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ex: <http://example.org/lovecraft#>

SELECT ?organisation ?member
WHERE {
    ?organisation rdf:type ex:Organisation .
    ?organisation ex:hasNotableMember0 ?member .
}
"""

print(query_str)
```

	Organisation	Member
0	League_of_Light	Ash_Williams
1	League_of_Light	Cybil
2	Brotherhood_of_the_Yellow_Sign	Chemb
3	Brotherhood_of_the_Yellow_Sign	Claret
4	Brotherhood_of_the_Yellow_Sign	Colia Thalna
...
73	Midnight_Sons	Ghost Rider
74	Carlyle_Expedition	Roger Carlyle
75	Brotherhood_of_the_Black_Pharaoh	Audrey Penhew the Dark Sorcerer
76	Brotherhood_of_the_Black_Pharaoh	Children of the Sphinx
77	Brotherhood_of_the_Black_Pharaoh	The Brotherhood cultists who operates in the d...

78 rows x 2 columns

	Organisation	MemberCount
0	Arkham_Detective_Agency_%28organisation%29	1
1	Avengers	3
2	Brotherhood_of_the_Black_Pharaoh	3
3	Brotherhood_of_the_Skin	9
4	Brotherhood_of_the_Yellow_Sign	20
5	Carlyle_Expedition	1
6	Children_of_Arion	1
7	Church_of_Starry_Wisdom	1
8	Cthulhu_Cult	1
9	Cult_of_Glaaki	1
10	Cult_of_the_Darkhold	1
11	Danger_Girl	3
12	Darkhold_Redemers	1
13	Esoteric_Order_of_Dagon	1
14	Ghostbusters	4
15	Great_Intelligence_Institute	1
16	League_of_Light	2
17	Midnight_Sons	2
18	S.M.A.R.T	1
19	Silver_Twilight_Lodge	5
20	Wilmarth_Foundation	3
21	X-Men	13

SPARQL-запросы

- На каких языках, кроме английского, были написаны произведения, связанные с мифологией Лавкрафта? Сколько произведений существует на каждом из этих языков? Какой язык является лидером по количеству произведений в текущей онтологии?

```
query_str = """
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ex: <http://example.org/lovecraft#>

SELECT ?work ?language ?author
WHERE {
  ?work rdf:type ex:Work .
  ?work ex:isInLanguageW ?language .
  ?work ex:hasAuthorW ?author .
  FILTER(?language != "English")
}
"""

```

Список произведений, их языков и авторов (отсортировано по языку):

	Work	Language	Author
0	Mother_of_Toads	Dutch	Clark_Ashton_Smith
1	The_Coming_of_the_White_Worm	Dutch	Clark_Ashton_Smith
2	The_Door_to_Saturn	Dutch	Clark_Ashton_Smith
3	Jerusalems_Lot	Dutch	Stephen_King
4	Vulthoom_%28short_story%29	Dutch	Clark_Ashton_Smith
...
211	Jerusalems_Lot	Spanish	Stephen_King
212	The_Cthulhu_Cycle	Spanish	Various
213	The_Dunwich_Horror_%281979_comic%29	Spanish	H._P._Lovecraft
214	Los_mitos_de_Cthulhu	Spanish	H._P._Lovecraft
215	Swedish_Cults	translated from Swedish	Anders_Fager

216 rows x 3 columns

Сводка по языкам (упорядочено от большего к меньшему):

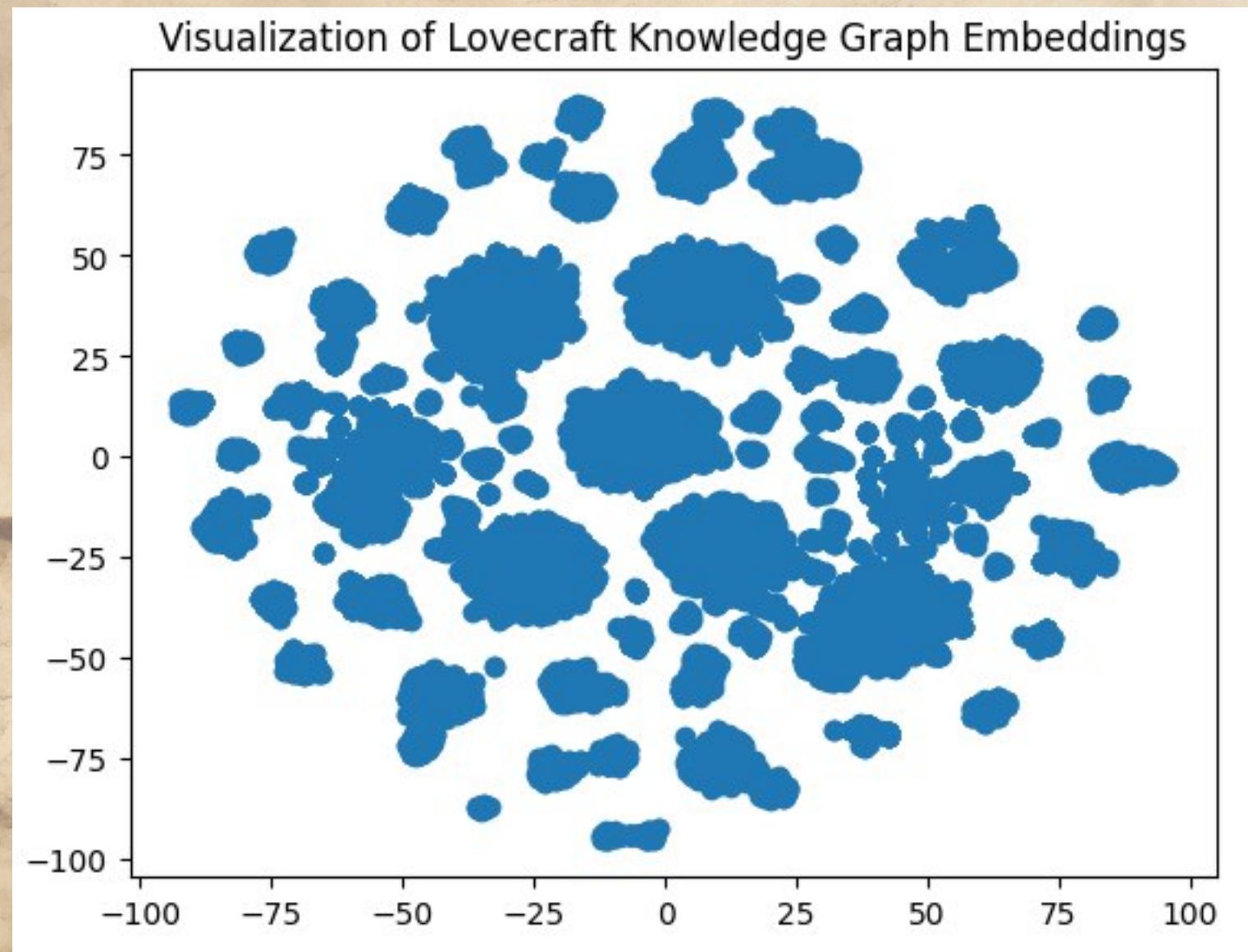
Language	Count
French	68
German	60
Italian	22
Spanish	19
Japanese	14
Portuguese	8
Dutch	7
Finnish	4
Polish	4
Hungarian	3
Russian	2
English French	1
Greek	1
Japanese (translated into English by Jim Rion)	1
Norwegian	1
translated from Swedish	1

Проблемы и вызовы

Проблема	Описание	Решение
Автоматизация парсинга	Обработка 18 страниц (~350 статей на каждой). Класс Unclassified составил 66 из 5400 статей.	Настроен парсер с классификацией (точность 99%), проведена первичная очистка и определены свойства графа.
Ошибки в данных	Некорректные форматы и символы, различные написания фамилии Лавкрафта.	Дополнительная валидация и экранирование.
Объемные аннотации	Описание сущностей (например, Ктулху, Лавкрафт) превышало 30 тыс. символов.	Использование модели T5 для генерации кратких описаний, сохранение исходных данных.
Обработка самоссылок	Некоторые сущности были и концептами, и индивидами (например, The Arkham Detective как произведение и концепт).	Особая обработка для корректного отображения двух уровней сущности.
Оптимизация графа	Избыточные данные и связи снижали производительность reasoning.	Удаление избыточных данных и оптимизация структуры графа.
Унификация данных	Данные из различных источников имели разные форматы.	Приведение данных к единому формату.

Эмбеддинги

1089/1089 [=====] - 118s 109ms/step
Mean Reciprocal Rank (MRR): 0.32
Mean Rank (MR): 1188.77
Hits@10: 0.43
Hits@3: 0.35
Hits@1: 0.27



Топ предсказания для произведений Лавкрафта:

	Work	Predicted Author
0	A_Confession_of_Unfaith	H.P. Lovecraft
1	A_Description_of_the_Town_of_Quebeck,_in_New_F...	H.P. Lovecraft
2	A_Reminiscence_of_Dr._Samuel_Johnson	H.P. Lovecraft
3	A_Winter_Wish_and_Other_Poems	H.P. Lovecraft
4	Ad_Criticos:_Liber_Secundus	H.P. Lovecraft
5	Advice_for_the_Weird_Fictioneer	H.P. Lovecraft
6	Against_Religion	H.P. Lovecraft
7	Alfredo:_A_Tragedy	H.P. Lovecraft
8	Amateur_Journalism:_Its_Possible_Needs_and_Bet...	H.P. Lovecraft
9	An_American_to_Mother_England	H.P. Lovecraft

	Score	Probability
0	4.184221	0.984994
1	1.648891	0.838741
2	2.120203	0.892851
3	1.729219	0.849312
4	3.416646	0.968221
5	3.111109	0.957349
6	3.700461	0.975884
7	1.934421	0.873738
8	3.772384	0.977520
9	3.882388	0.979814

Accuracy: 100.00%

Classification Report:				
	precision	recall	f1-score	support
H.P. Lovecraft	1.00	1.00	1.00	268
accuracy				268
macro avg	1.00	1.00	1.00	268
weighted avg	1.00	1.00	1.00	268

Инсайт

Топ-3 альтернативных автора:	
Brian Lumley	153
Thomas Ligotti	103
Frank Belknap Long	7



Спасибо за внимание