

HIMIKO：歴史史料記述を知識グラフとして 表現する試み

小川 潤

(ROIS-DS人文学オープンデータ共同利用センター)
第1回ヒストリーテック勉強会 2023年11月22日

アノテーションの目的

- 史料そのものの来歴や性質といった情報（メタデータ）に留まらず、**その中にどのような事柄が実際に記されているのか**を、機械可読なデータとして構造化する必要がある
 - > より「深い」分析に活用可能なデータの作成
- 専門の歴史研究者が、（比較的）手軽に、**日常の史料読解作業の延長で「構造化された」「共有可能な」データを生み出せる環境**を作る
 - > 情報の粒度と、ある程度の汎用性が求められる
 - > 解釈を前提とした構造化

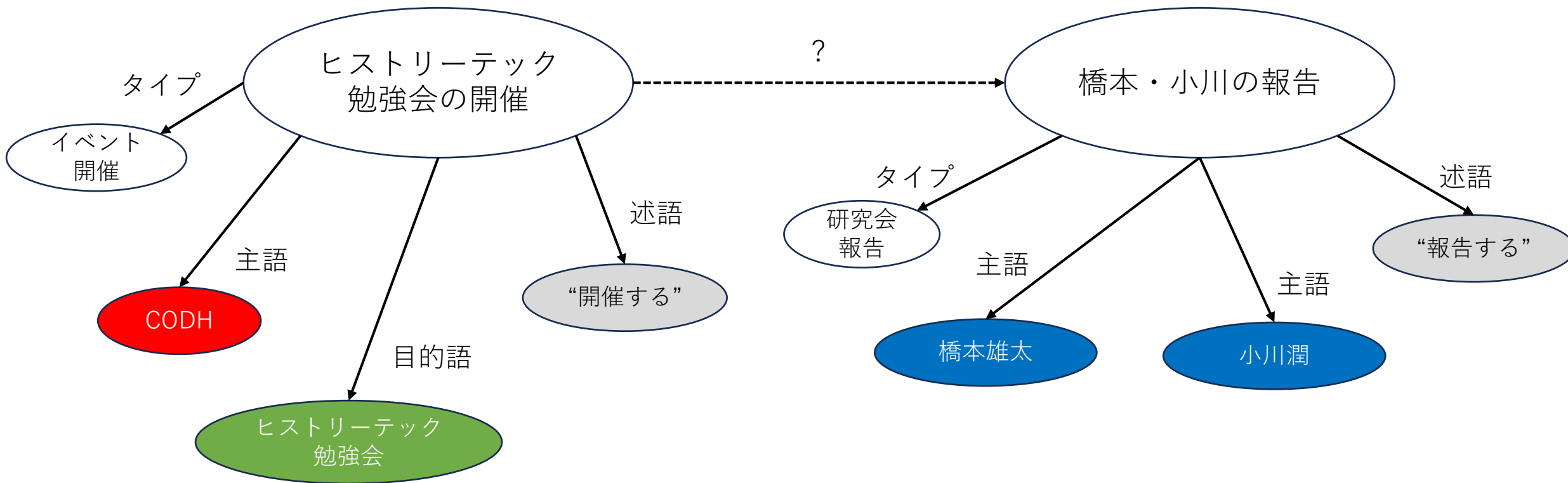
どのようにアノテーションするか

- RDFを用いて、**知識グラフ**として史料記述の構造化を行う
 - 史料中のあらゆる情報項目にURIを与える
 - 主語・述語・目的語のトリプルであらゆる情報を表現する
 - 典拠識別子（できればLOD）ともリンクする
- 既存のイベントオントロジーやFactoidモデルを参照する
 - 「出来事」に関わる情報をデータとして記述する方法論
 - 人・場所・時間といった情報をどのように組織化するか

歴史マイクロナレッジとHIMIKOオントロジー

- 出来事や状況に言及する「個別の記述」を単位にデータ構造化
 - 出来事や状況に関する記述自体をデータリソースとする
 - 個々の記述箇所を独立したデータ項目として扱う
 - reification: 本来は述語として記述される「関係性」自体を主語として記述
 - 史料を「断片化」してデータ構造化
 - 断片化された記述単位でメタデータや属性情報を付与したり、
 - 他のデータとリンクしたり共有したりすることで、
 - ミクロなレベルでの横断的知識接続が可能に

――。2023年11月22日、CODHは第一回ヒストリーテック勉強会を開催し、橋本雄太と小川潤が報告した。――。



- ざっくりデータ (Turtle) で表すと、

```
<http://www.example.com/ヒストリーテック勉強会の開催>  
  a <http://www.example.com/イベント開催>;  
  <http://www.example.com/主語> <http://www.example.com/CODH>;  
  <http://www.example.com/目的語> <http://www.example.com/ヒストリーテック勉強会>;  
  <http://www.example.com/述語> “開催する”.
```

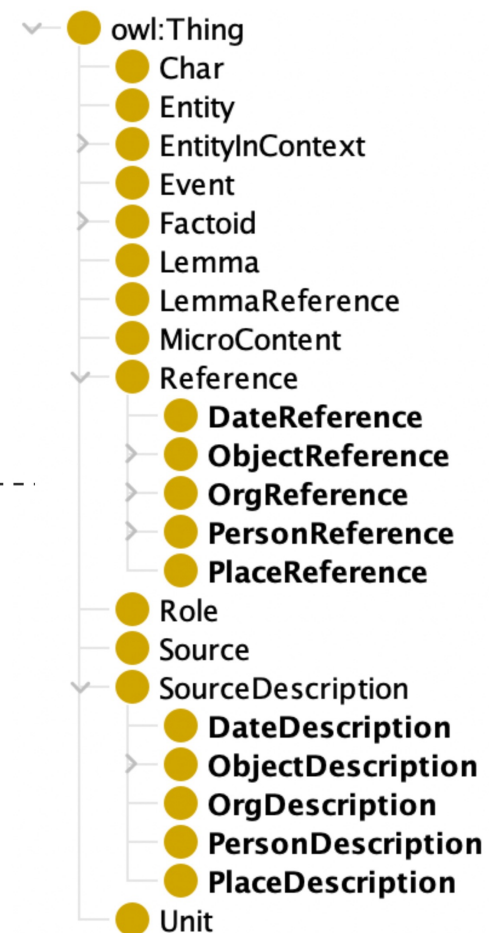
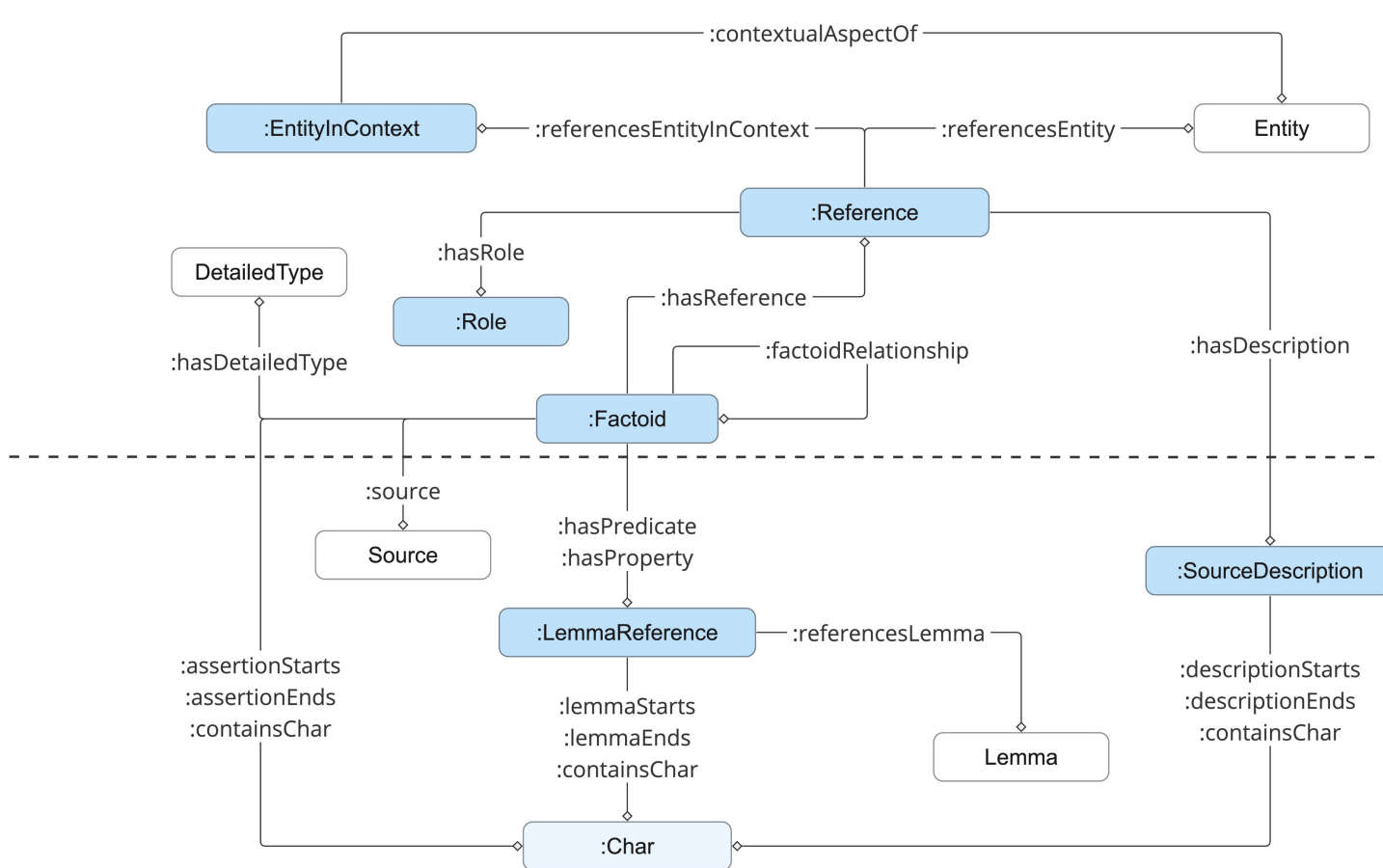
```
<http://www.example.com/橋本・小川の報告>  
  a <http://www.example.com/勉強会報告>;  
  <http://www.example.com/主語> <http://www.example.com/橋本雄太>;  
  <http://www.example.com/主語> <http://www.example.com/小川潤>;  
  <http://www.example.com/述語> “報告する”.
```

こうしたデータを作っていくためのルールとシステムが必要！

歴史マイクロナレッジとHIMIKOオントロジー

- 出来事や状況に言及する「個別の記述」を単位にデータ構造化
 - 出来事や状況に関する記述自体をデータリソースとする
 - * reification
 - 個々の記述ごとにメタデータや属性情報を付与する
 - 史料を断片化してデータ構造化
 - 断片化された記述単位でリンクしたり、共有したり
- **HIMIKO** (Historical Micro Knowledge and Ontology)
 - 歴史マイクロナレッジを実装するオントロジー

知識の構造化



史料記述の構造化

- 特徴

- Referenceの記述

[Factoidモデル](#)において導入された概念で、「特定のFactoidにおいて言及されるエンティティ」を表す

-> かなり**contextualな情報**であり、特定の文脈におけるエンティティの「役割」などが記述される

- エンティティの表記や文字情報を含む高い一次史料参照性

SourceDescriptionやCharといった項目はいずれも**一次史料の具体的な記述に準拠したデータ**を記述するためのもの

HIMIKOエディタ

- HIMIKOオントロジーに基づく知識グラフ構築を支援するエディタ

- 基本情報

- Nuxt JS + Vuetify
- Firebase
- Vercel
- <https://himiko-editor.vercel.app/>

- 機能

- エンティティマークアップ
- エンティティリンクング
- **知識グラフ編集**



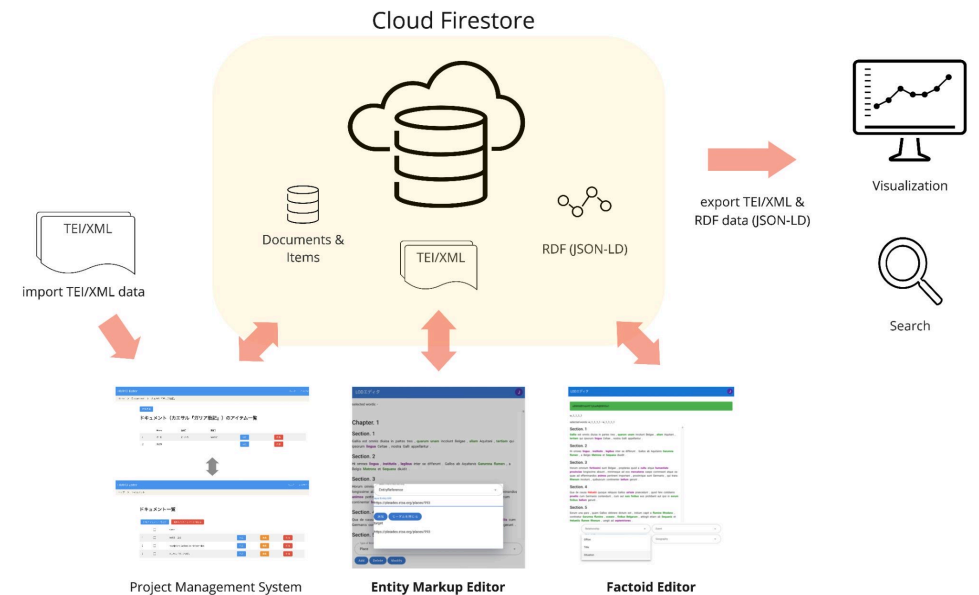
TEIテキストに対するマークアップとLOD化を支援するツールです。

[プロジェクトをみる](#)

お知らせ [すべて](#)

2022-12-06 Under Construction.

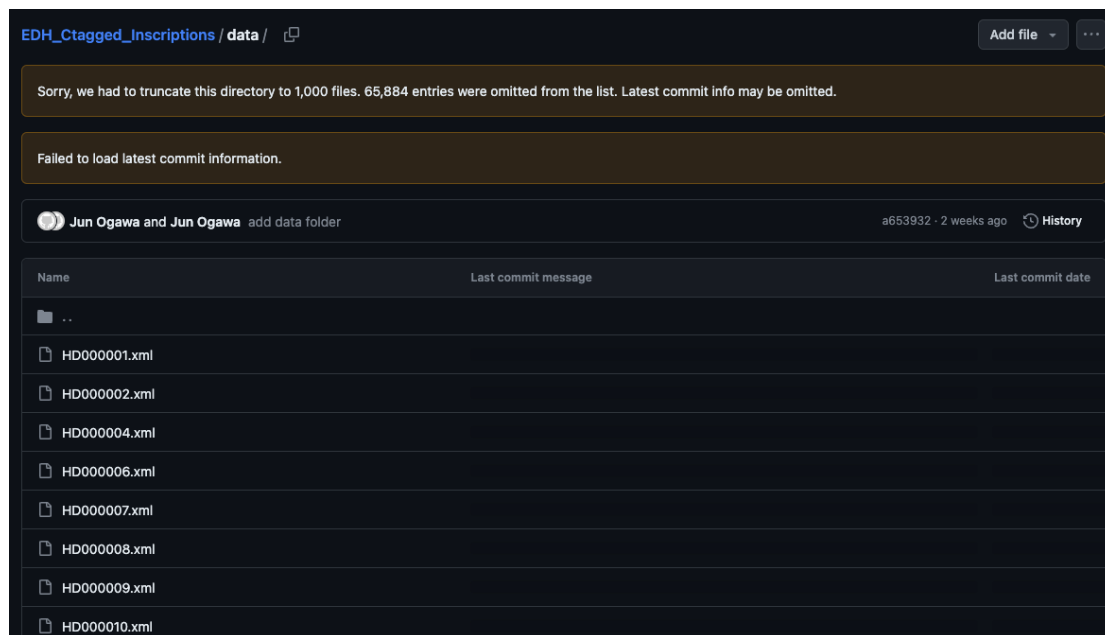
2023-01-24 ニュースのタイトル



- データのアップロード

- テキストデータのアップロード

- <c>タグで文字レベルの構造化がなされた TEI/XML ファイル
 - 文字レベルでのURI付与を実現するため



```
▼<line n="1">
  ▼<expan>
    ▼<abbr>
      <c xml:id="L1C1">D</c>
    </abbr>
    ▼<ex>
      <c xml:id="L1C2">i</c>
      <c xml:id="L1C3">s</c>
    </ex>
  </expan>
  <c xml:id="L1C4"> </c>
  ▼<expan>
    ▼<abbr>
      <c xml:id="L1C5">M</c>
    </abbr>
    ▼<ex>
      <c xml:id="L1C6">a</c>
      <c xml:id="L1C7">n</c>
      <c xml:id="L1C8">i</c>
      <c xml:id="L1C9">b</c>
      <c xml:id="L1C10">u</c>
      <c xml:id="L1C11">s</c>
    </ex>
  </expan>
  <c xml:id="L1C12"> </c>
</line>
```

テキスト貼付でアップロード

URL指定でアップロード

HIMIKO Editor

ライトモード 日本語

参加しているプロジェクト プロジェクト ドキュメント アイテム

TEIファイルのアップロード

upload **paste** **url**

XML URL
https://junjun7613.github.io/EDH_Ctagged_Inscriptions/data/HD000435.xml

Target Element
surface

Target Attribute
type

Target Attribute Value
text

create

閉じる

インターフェイスに表示する
テキストを含む要素（エレメ
ント）名を指定

属性（アトリビュート）情報
による絞り込み

アイテム詳細

user_id	HgURY4ZRkwQy2fadN7Gc2IZfi8w1
project_id	NO7r18txVG04OcfjCmFq
document_id	bn8T6JYTPysljHF7GOgw
item_id	438KV8NEvh8TQCBPnMVN

[editor](#) [TEIファイルのアップロード](#) [\(構築中\) IIIIFマニフェストのアップロード](#) [ダウンロード](#)

↑
エディタ表示

HIMIKO Editor

エンティティ ファクトイド

👤 📍 📅 🗉 ❤️

選択範囲

テキスト

Garumna

開始ID

http://www.example.com/roman-ontology/resource/438KV8NEvh8TQCBPnMVN#1_1_225

終了ID

http://www.example.com/roman-ontology/resource/438KV8NEvh8TQCBPnMVN#1_1_231

文字（ここでは‘G’）を表すURI

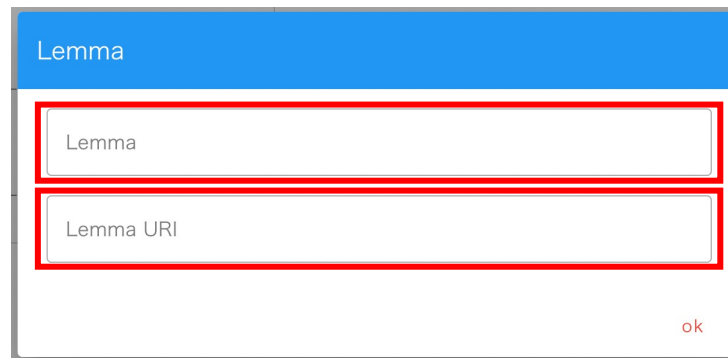
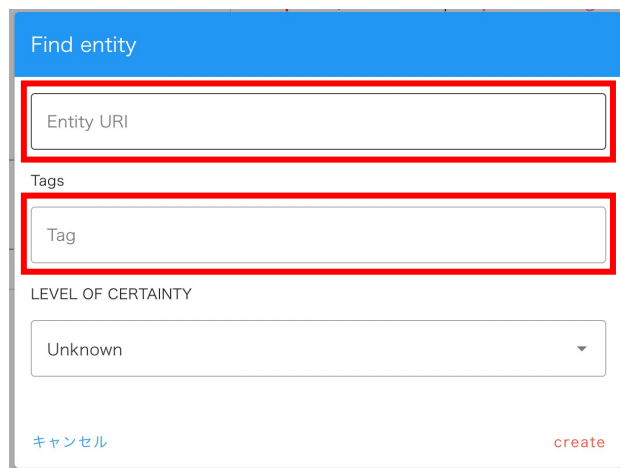


Gallia est omnis divisa in partes
aliam Aquitani, tertiam qui ipsori
lantur. Hi omnes lingua, instituti:
Aquitanis Garumna flumen, a Belg
Horum omnium fortissimi sunt Be
humanitate provinciae longissime a
saepe commeant atque ea quae
important, proximique sunt Germ
quibuscum continenter bellum geru
reliquos Gallos virtute praecedunt,
Germanis contendunt, cum aut su
eorum finibus bellum gerunt. Eo
dictum est, initium capit a flumin
flumina Oceano finibus Belgarum

• エンティティのマークアップとリンク

外部典拠データとのリンク

タグの付与



辞書見出し形の入力

語彙典拠データへのリンク



画像とリンクする

Gallia est omnis divisa in partes tres, quarum **unam** incolunt **Belgae**, **aliam** **Aquitani**, **tertiam** qui **ipsorum lingua** **Celtae**, **nostra Galli** appellantur. Hi omnes lingua, institutis, legibus inter se differunt. Gallos ab Aquitanis Garumna flumen, a **Belgis** Matrona et Sequana dividit. Horum omnium fortissimi sunt Belgae, propterea quod a cultu atque humanitate provinciae longissime absunt,

エンティティ ファクトイド

終了ID
http://www.example.com/roman-ontology/resource/438KV8NEvh8TQCBPnMVN#1_1_6

選択されたEntity

URI
<http://www.example.com/roman-ontology/resource/Description/e85961e7-d437-4d06-a988-a0662e85d95a>

type
 PlaceDescription

text
 Gallia

reference
<https://pleiades.stoa.org/places/993>

tags
 region

更新 削除

Gallia est omnis divisa in quitani, tertiam qui ipsorui omnes lingua, institutis, legi flumen, a Belgis Matrona e Belgae, propterea quod a c minimeque ad eos mercator animos pertinent important, incolunt, quibuscum continen reliquos Gallos virtute praec contendunt, cum aut suis f bellum gerunt. Eorum una capit a flumine Rhodano, c attingit etiam ab Sequanis Belgae ab extremis Galliae fluminis Rheni, spectant in Garumna flumine ad Pyrena Hispaniam pertinet; spectat

438KV8NEvh8TQCBPnMVN	entities	1befcba1-17bc-42b7-80de-f5e88189faa8
+ コレクションを開始	+ ドキュメントの追加	+ コレクションを開始
entities	1befcba1-17bc-42b7-80de-f5e...	+ フィールドを追加
lemmas	22fd98b5-066d-4280-9834-c3f...	created_at: 2023年11月20日 17:18:12 UTC+9
+ フィールドを追加	3a6c770b-e58e-4504-9546-148...	data
targetAttribute: "type"	509aea5a-8835-4f46-83e6-5a6...	@id: "http://www.example.com/roman-ontology/resource/Description/1befcba1-17bc-42b7-80de-f5e88189faa8"
targetAttributeValue: "text"	767aedd-9911-4584-b8d7-27a...	@type: "https://junjun7613.github.io/MicroKnowledge/himiko.owl#Org"
targetElement: "div"	78fadd65-7878-4df7-9fed-8dc...	http://www.w3.org/2000/01/rdf-schema#label: "Aquitani"
title: "1-1"	856a9266-deb1-445c-a2a7-f27...	https://junjun7613.github.io/MicroKnowLedge/himiko.owl#
updated: 2023年11月20日 17:10:32 UTC+9	b2207dad-0dc4-4bd0-93b6-913...	0 "http://www.example.com/roman-ontology/resource/438KV8NEvh8TQCBPnMVN#1_1_76"
uri: "http://www.example.com/roman-ontology/resource/438KV8NEvh8TQCBPnMVN#1_1_6"	dc7c6878-a603-486a-8f1c-959...	1 "http://www.example.com/roman-ontology/resource/438KV8NEvh8TQCBPnMVN#1_1_77"
xml_path: "xml/438KV8NEvh8TQCBPnMVN#1_1_6"	e85961e7-d437-4d06-a988-a06...	2 "http://www.example.com/roman-ontology/resource/438KV8NEvh8TQCBPnMVN#1_1_78"

エンティティ ファクトイド

終了ID
http://www.example.com/roman-ontology/resource/438KV8NEvh8TQCBPnMVN#1_1_56

選択されたLemma

URI
<http://www.example.com/roman-ontology/resource/Lemma/62d07066-ec31-4898-aed9-2c47948cbf5d>

reference
<https://lila-erc.eu/data/id/lemma/107025>

text
 incolunt

更新 削除

Gallia est omnis divisa in partes tres, quarum unam incolunt quitani, tertiam qui ipsorum lingua Celtae, nostra Galli appe omnes lingua, institutis, legibus inter se differunt. Gallos ab Aq flumen, a Belgis Matrona et Sequana dividit. Horum omnium Belgae, propterea quod a cultu atque humanitate provinciae lor minimeque ad eos mercatores saepe commeant atque ea quae animos pertinent important, proximique sunt Germanis, qui tra incolunt, quibuscum continenter bellum gerunt. Qua de causa l reliquos Gallos virtute praecedunt, quod fere cotidianis proeliis contendunt, cum aut suis finibus eos prohibent aut ipsi in eo bellum gerunt. Eorum una pars, quam Gallos obtinere dictum capit a flumine Rhodano, continetur Garumna flumine, Oceano, attingit etiam ab Sequanis et Helvetiis flumen Rhenum, vergit ; Belgae ab extremis Galliae finibus oriuntur, pertinent ad inferior fluminis Rheni, spectant in septentrionem et orientem solem. Garumna flumine ad Pyrenaeos montes et eam partem Oceani Hispaniam pertinet; spectat inter occasum solis et septentriones.

• 知識グラフの編集

ファクトイドを作成

Gallia est omnis divisa in partes tres, quarum **unam** incolunt **Belgae**, **aliam** **Aquitani**, **tertiam** qui **ipsorum lingua Celtae**, **nostra Galli** appellantur. Hi omnes lingua, institutis, legibus inter se differunt. Gallos ab Aquitanis Garumna flumen, a Belgis Matrona et Sequana dividit. **Horum omnium fortissimi sunt Belgae, propterea quod a cultu atque humanitate provinciae longissime absunt**, minimeque ad eos mercatores saepe commeant atque ea quae ad effeminandos animos pertinent important, proximique sunt Germanis, qui trans Rhenum incolunt, quibuscum continenter bellum gerunt. Qua de causa Helvetii quoque reliquos Gallos virtute praecedunt, quod fere cotidianis proeliis cum Germanis contendunt, cum aut suis finibus eos prohibent aut ipsi in eorum finibus bellum gerunt. Eorum una pars, quam Gallos obtinere dictum est, initium capit a flumine Rhodano, continetur Garumna flumine, Oceano, finibus Belgarum, attingit etiam ab Sequanis et Helvetiis flumen Rhenum, vergit ad septentriones. Belgae ab extremis Galliae finibus oriuntur, pertinent ad inferiorem partem fluminis Rheni, spectant in septentrionem et orientem solem. Aquitania a Garumna flumine ad Pyrenaeos montes et eam partem Oceani quae est ad Hispaniam pertinet; spectat inter occasum solis et septentriones.

知識グラフ編集画面

データベースに登録

close

テキスト範囲を決定する

Horum omnium fortissimi sunt Belgae, propterea quod a cultu atque humanitate provinciae longissime absunt

Factoid Type

Select Factoid type ...

Factoidの詳細タイプを入力してください。

Factoidの説明を入力してください。

出来事や関係性の種類や詳細を入力

Factoid Type

Select Factoid type ...

Factoidの詳細タイプを入力してください。

Factoidの説明を入力してください。

いずれも文字列で入力

Event

▶ Event

▼ State of Affairs

Situation

Office

Occupation

Possession

Title

Property

▼ Relationship

FamilialRelationship

エンティティをFactoidに紐付け

ファクトイドを作成

Gallia est omnis divisa in partes tres, quarum **una**
m incolunt **Belgae**, **aliam Aquitani**, **tertiam** qui **ips**
orum lingua Celtae, **nostra Galli** appellantur. Hi
omnes lingua, institutis, legibus inter se differunt.
Gallos ab Aquitanis Garumna flumen, a Belgis
Matrona et Sequana dividit. Horum omnium
fortissimi sunt Belgae, propterea quod a cultu atque
humanitate provinciae longissime absunt, minimeque
ad eos mercatores saepe commeant atque ea quae
ad effeminandos animos pertinent important,
proximique sunt Germanis, qui trans Rhenum
incolunt, quibuscum continenter bellum gerunt. Qua
de causa Helvetii quoque reliquos Gallos virtute
praecedunt, quod fere cotidianis proeliis cum
Germanis contendunt, cum aut suis finibus eos

該当するエンティティ
をクリック

Has Reference

Referenceを選択してください。

Subject

カードを追加

<https://junjun7613.github.io/MicroKnowledge/himiko.owl#subject>

Optional: EntityInContext URI

Roleを入力してください。

Optional: Unitを入力

Optional: Quantityを入力

選択したEntityを一時保存

どのようなセマンティック
でエンティティをFactoid
に紐づけるかを選択

Factoidの文脈における
エンティティの「役割」
や数量等、文脈に依拠す
る情報を入力する

エンティティをFactoidに紐付け

ファクトイドを作成

Gallia est omnis divisa in partes tres, quarum **una**
m incolunt **Belgae**, **aliam Aquitani**, **tertiam** qui **ips**
orum lingua Celtae, **nostra Galli** appellantur. Hi
omnes lingua, institutis, legibus inter se differunt.
Gallos ab Aquitanis Garumna flumen, a Belgis
Matrona et Sequana dividit. Horum omnium
fortissimi sunt Belgae, propterea quod a cultu atque
humanitate provinciae longissime absunt, minimeque
ad eos mercatores saepe commeant atque ea quae
ad effeminandos animos pertinent important,
proximique sunt Germanis, qui trans Rhenum
incolunt, quibuscum continenter bellum gerunt. Qua
de causa Helvetii quoque reliquos Gallos virtute
praecedunt, quod fere cotidianis proeliis cum
Germanis contendunt, cum aut suis finibus eos

テキストインター
フェイス操作

Factoid relationship

Select Factoid Relationship Type ...

カードを追加

Factoid間の関係性を記述

Has Predicate/Property

Predicate/Propertyを選択してください。

カードを追加

出来事や関係性を述べる述語をテキストから選択・入力

データのダウンロード

アイテム詳細

user_id	HgURY4ZRkwQy2fadN7Gc2IZf8w1
project_id	NO7r18txVG04OcfjCmFq
document_id	bn8T6JYTPysljHF7GOgw
item_id	438KV8NEvh8TQCBPnMVN

editor | TEIファイルのアップロード | (構築中) IIIF Manifestのアップロード | **ダウンロード**

438KV8NEvh8TQCBPnMVN_entities.json	昨日 8:30	14 KB	JSON File
438KV8NEvh8TQCBPnMVN_entities.ttl	昨日 8:30	13 KB	テキスト…ディット
438KV8NEvh8TQCBPnMVN_lemmas.json	昨日 8:30	3 KB	JSON File
438KV8NEvh8TQCBPnMVN_lemmas.ttl	昨日 8:30	3 KB	テキスト…ディット
438KV8NEvh8TQCBPnMVN_lds.json	昨日 8:30	12 KB	JSON File
438KV8NEvh8TQCBPnMVN_lds.ttl	昨日 8:30	12 KB	テキスト…ディット
438KV8NEvh8TQCBPnMVN_references.json	昨日 8:30	6 KB	JSON File
438KV8NEvh8TQCBPnMVN_references.ttl	昨日 8:30	5 KB	テキスト…ディット
438KV8NEvh8TQCBPnMVN.xml	昨日 8:30	44 KB	XML Document

JSON-LDとTurtleでダウンロード

```
{  
  "https://junjun7613.github.io/MicroKnowledge/himiko.owl#hasNote": "",  
  "https://junjun7613.github.io/MicroKnowledge/himiko.owl#hadParticipant": [  
    "http://www.example.com/roman-ontology/resource/Reference/  
    22fd98b5-066d-4280-9834-c3ffa0a88d46_3ff4ef85-7809-49ef-a289-0fba7eaf404d",  
    "http://www.example.com/roman-ontology/resource/Reference/  
    dc7c6878-a603-486a-8f1c-959117f98a8b_3ff4ef85-7809-49ef-a289-0fba7eaf404d"  
  ],  
  "https://schema.org/dateModified": {  
    "@value": "2023-11-20",  
    "@type": "http://www.w3.org/2001/XMLSchema#date"  
  },  
  "@id": "http://www.example.com/roman-ontology/resource/3ff4ef85-7809-49ef-a289-0fba7eaf404d",  
  "https://junjun7613.github.io/MicroKnowledge/himiko.owl#tookPlaceAt": [  
    "http://www.example.com/roman-ontology/resource/Reference/  
    856a9266-deb1-445c-a2a7-f271d97e64fe_3ff4ef85-7809-49ef-a289-0fba7eaf404d"  
  ],  
  "https://junjun7613.github.io/MicroKnowledge/himiko.owl#hasDetailedType": "居住",  
  "https://junjun7613.github.io/MicroKnowledge/himiko.owl#hasPredicate": [  
    "http://www.example.com/roman-ontology/resource/Lemma/62d07066-ec31-4898-aed9-2c47948cbf5d"  
  ],  
  "@type": "https://junjun7613.github.io/MicroKnowledge/himiko.owl#SituationFactoid",  
  "https://schema.org/dateCreated": {  
    "@value": "2023-11-20",  
    "@type": "http://www.w3.org/2001/XMLSchema#date"  
  },  
  "https://junjun7613.github.io/MicroKnowledge/himiko.owl#source": "",  
  "https://schema.org/creator": {  
    "@id": "https://orcid.org/1985783-109"  
  },  
  "@context": "https://junjun7613.github.io/MicroKnowledge/himiko-context.jsonld"  
}
```

Factoidの作成者や作成日時に関する情報は自動で付与される

データの利用

「居住」に関する記述と、そこで言及されている地名を検索するクエリと出力結果

```

1 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
2 select * where {
3   ?s <https://junjun7613.github.io/MicroKnowledge/himiko.owl#hasDetailedType> "居住"^^<xsd:string>.
4   ?s <https://junjun7613.github.io/MicroKnowledge/himiko.owl#tookPlaceAt> ?o.
5 } limit 100
6

```

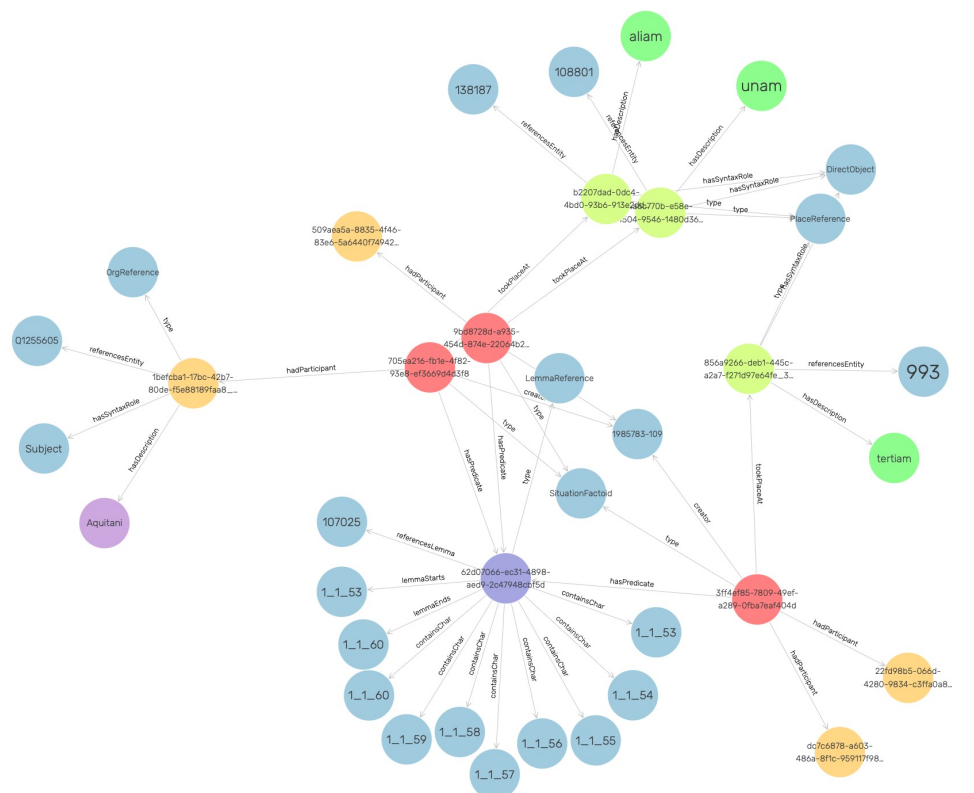
Press Alt+Enter to autocomplete

Table Raw Response Pivot Table Google Chart Download as

Filter query results Showing results from 1 to 3 of 3. Query took 0.2s, moments ago.

	s	o
1	http://www.example.com/roman-ontology/resource/3ff4ef85-7809-49ef-a289-0fba7eaf404d	http://www.example.com/roman-ontology/resource/Reference/856a9266-deb1-445c-a2a7-f271d97e64fe_3ff4ef85-7809-49ef-a289-0fba7eaf404d
2	http://www.example.com/roman-ontology/resource/705ea216-fb1e-4f82-93e8-ef3669d4d3f8	http://www.example.com/roman-ontology/resource/Reference/b2207dad-0dc4-4bd0-93b6-913e2dd43d6e_705ea216-fb1e-4f82-93e8-ef3669d4d3f8
3	http://www.example.com/roman-ontology/resource/9bd8728d-a935-454d-874e-22064b2c49d7	http://www.example.com/roman-ontology/resource/Reference/3a6c770b-e58e-4504-9546-1480d362f010_9bd8728d-a935-454d-874e-22064b2c49d7

検索



可視化
(グラフDB)

今後の開発の展望

- アノテーションに基づく可視化の実装
 - エンティティマークアップに基づくマッピングなど
 - Factoid記述に基づく**史料内容の「深い」探索・分析**
 - ・意味ネットワークの可視化や推論など
- テキスト、画像、さらには3Dの紐付け
 - IIF Content State APIを用いた画像Factoidエディタ（開発中）
 - テキスト情報を3D空間・オブジェクト上に表示するビューア（こちらも開発中）
 - これらを統合することを目指す

論点

- いくつかのジレンマ
 - 特殊性 × 汎用性
 - 精密性 × 簡便性
 - 柔軟性 × 共有可能性
- 典拠識別子が十分に整備されていない分野はどうする？
 - 識別子に限らず、テキストなどの基礎データ全般
- 自動化はどこに、どの程度取り入れていくべきか、いくことができるか？