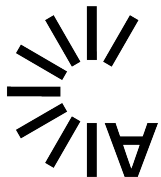


*

Some words and phrases on the slides may be shocking or offensive. They are provided solely as an illustration and explanation of the research results and do not reflect the opinions of the speakers or their organisations.



Cultural AI
a lab for
culturally
valued AI



Centrum Wiskunde & Informatica



Culturally Aware AI

Uncovering and modelling cultural controversies in data

Andrei Nesterov, PhD student at CWI / Human-Centered Data Analytics

Promotor: Jacco van Ossenbruggen, VU

Co-promotor: Laura Hollink, CWI

Co-promotor: Marieke van Erp, KNAW

Scientific meeting at CWI, 30 September 2022

Digital cultural heritage collections are vast



650 K

objects with
metadata

Rijksmuseum



17 M

digitised pages of
historical newspapers

KB National Library

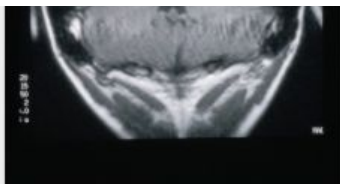


54 M

searchable
items

Europeana

Controversial language in object descriptions

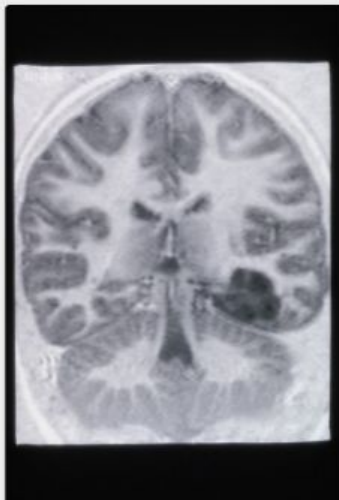


MRI scan; brain cancer
(primitive neuroectodermal)

Wellcome Collection

Wellcome museum, primitive
medicine: hygiene

Wellcome Collection



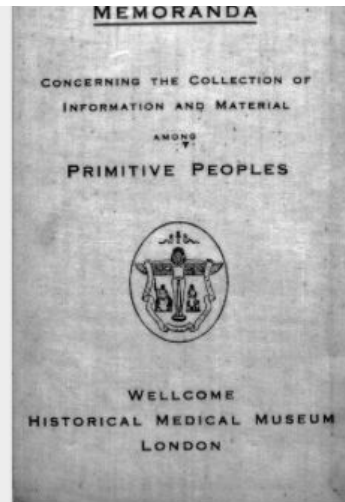
MRI scan; brain cancer
(primitive neuroectodermal)

Wellcome Collection



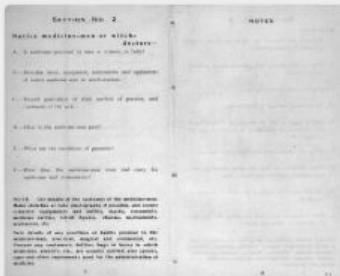
Wellcome museum, primitive
medicine: hygiene

Wellcome Collection



Memoranda...primitive
peoples

Wellcome Collection



Memoranda...primitive
peoples

Wellcome Collection



A Bochev-Dohrmann-
Gunzburger stabilization



Menu **nrc** Mijn nieuws

O
opinie

Schrap de term 'Bersiap' want die is racistisch

Dekolonisatie De term 'Bersiap' is te generaliserend om te gebruiken als naam voor de periode na de Japanse capitulatie. *Triyana* heeft de term daardoor een racistische lading.

10 januari 2022
Leestijd 4 minuten
Opslaan in leeslijst



NOS Nieuws Sport Uitzendingen

NOS NIEUWS • BINNENLAND • DI 11 JANUARI, 02:23

'Bersiap' niet gebruikt op Indonesië-tentoonstelling Rijksmuseum



Contentious terms in museum databases



"Aboriginal"
"Berber"
"Black"
"Causasian"
"Disabled"
"Exotic"
"Gypsy"
"Half-blood"
"Indian"
"Native"
"Primitive"
"Trans"
"Western"

...
55 terms

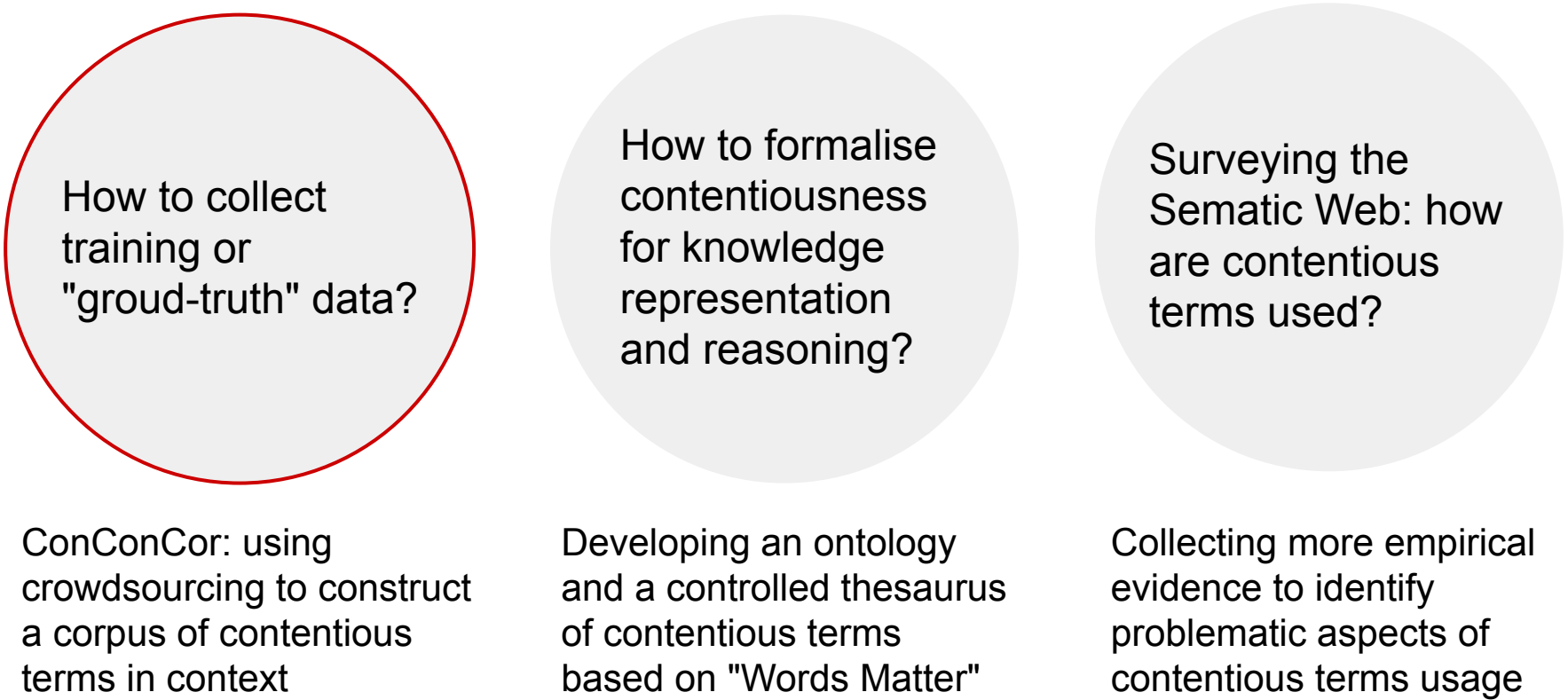
(1) Museum Volkenkunde. Photo Peter Hilz

(2) Modest, Wayne & Lelijveld, Robin (editors) 2018. Words Matter, Work in Progress I. National Museum of World Cultures.

<https://www.materialculture.nl/en/publications/words-matter>

How can we develop and use computational approaches that capture, process, and represent socio-cultural sensitivities to make digital cultural heritage more inclusive?

Computational approaches to contentious terms



How to collect training or "groud-truth" data?

ConConCor: using crowdsourcing to construct a corpus of contentious terms in context

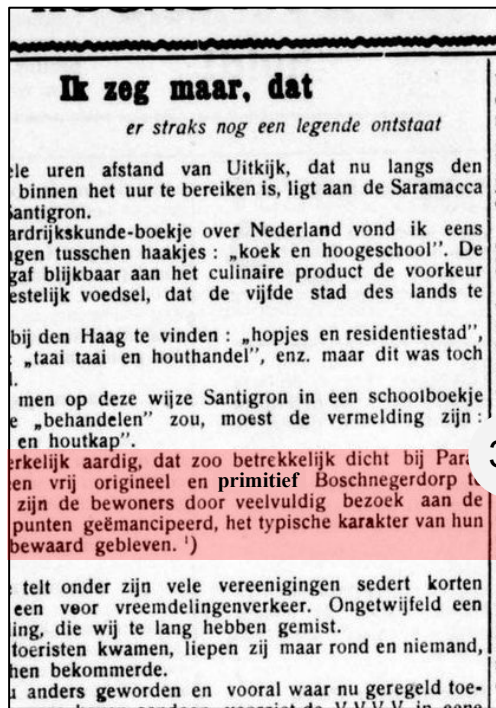
How to formalise contentiousness for knowledge representation and reasoning?

Developing an ontology and a controlled thesaurus of contentious terms based on "Words Matter"

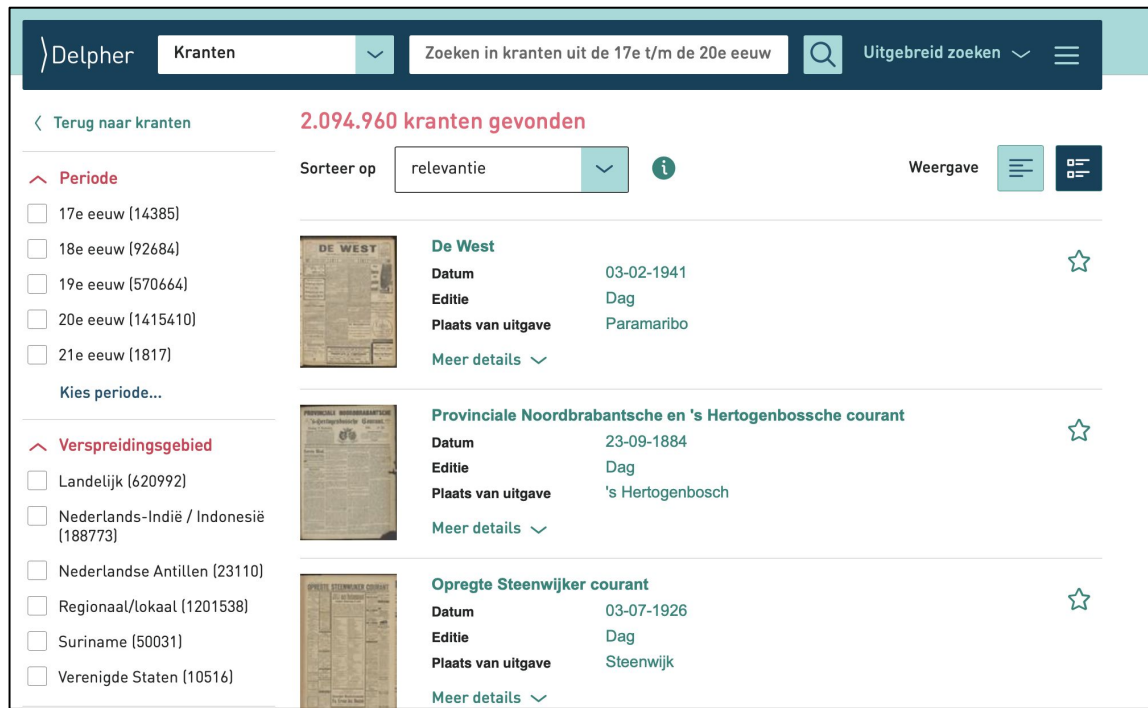
Surveying the Sematic Web: how are contentious terms used?

Collecting more empirical evidence to identify problematic aspects of contentious terms usage

Building a training corpus

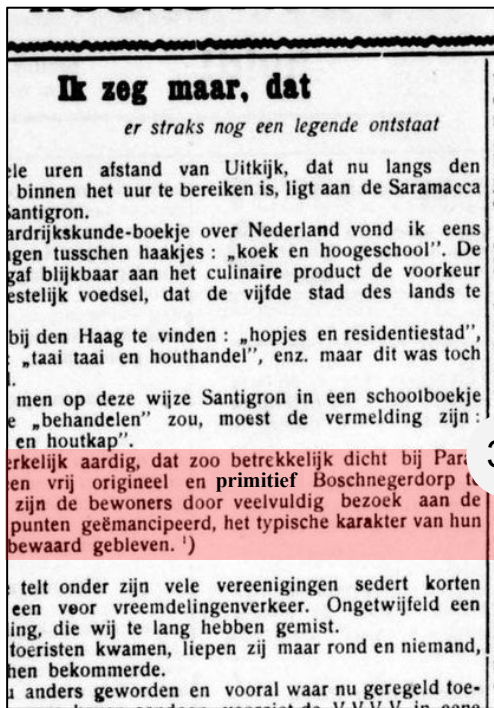


Screenshot. Delpher. KB National Library of the Netherlands:
<https://www.delpher.nl/nl/kranten/view?coll=ddd&identifier=ddd:110634579:mpeg21:a0062>



Screenshot. Delpher. <https://www.delpher.nl/nl/kranten/results?query=&coll=ddd&titel>

Crowdsourcing for autodetection of contentiousness?



Screenshot. Delpher. KB National Library of the Netherlands:
<https://www.delpher.nl/nl/kranten/view?coll=ddd&identifier=ddd:110634579:mpeg21:a0062>

Constructing an annotated Dutch corpus of contentious terms in contexts (ConConCor)

+ Annotator responses:

- *Contentious*
- *Non-contentious*
- *I don't know*
- *Bad OCR*

+ Newspapers metadata

+ Annotators demographics



> 400
laypeople
and
experts

7
annotators
per sample

Ryan Brate, Andrei Nesterov, Valentin Vogelmann, Jacco van Ossenbruggen, Laura Hollink, and Marieke van Erp. 2021. Capturing Contentiousness: Constructing the Contentious Terms in Context Corpus. In Proceedings of the 11th on Knowledge Capture Conference (K-CAP '21). Association for Computing Machinery, New York, NY, USA, 17–24.
DOI: <https://doi.org/10.1145/3460210.3493553>

Contentiousness by term: majority agreement

non-contentious

zimbabwe
trans
ontdekken
mohammedaan
kaukasisch
jakarta
indisch
birma
bediende
batavia
allochtoon

dwerg
zigeunermuziek
kaffer
eskimo
misvormden
slaaf
mulat
hottentot
slaven
pygmee
exotisch

...

52 terms

contentious

negerslaven
negerrijk
negercultuur
neger
bosneger
aapman



In all the given snippets, these terms
are contentious according to
majority of annotators

Crowdsourcing contentiousness: main results

1. Context is a deciding factor in the annotators' judgments of contentiousness

For example, **220** annotators each individually annotated at least 1 term both as contentious and non-contentious across different samples

2. Despite **low overall inter-annotator agreement**, **42%** of the samples in the corpus is annotated with a **high percentage agreement (81–100%)** between annotators

Median Krippendorff's α : **0,39** among laypeople annotators; **0,53** – experts ($-1 < K\alpha < 1$)
(calculated for every batch of annotated samples)

Experiments with ConConCor and future work

- How to capture contentiousness of a term taking its context into account?

An experiment with Linear regression model showed that word embeddings may capture terms' semantic which is indicative of their contentiousness

- Can we detect new (unknown) contentious words or phrases?



Ryan Brate
PhD student
KNAW HuC DHLab



Computational approaches to contentious terms

How to collect training or "ground-truth" data?

ConConCor: using crowdsourcing to construct a corpus of contentious terms in context

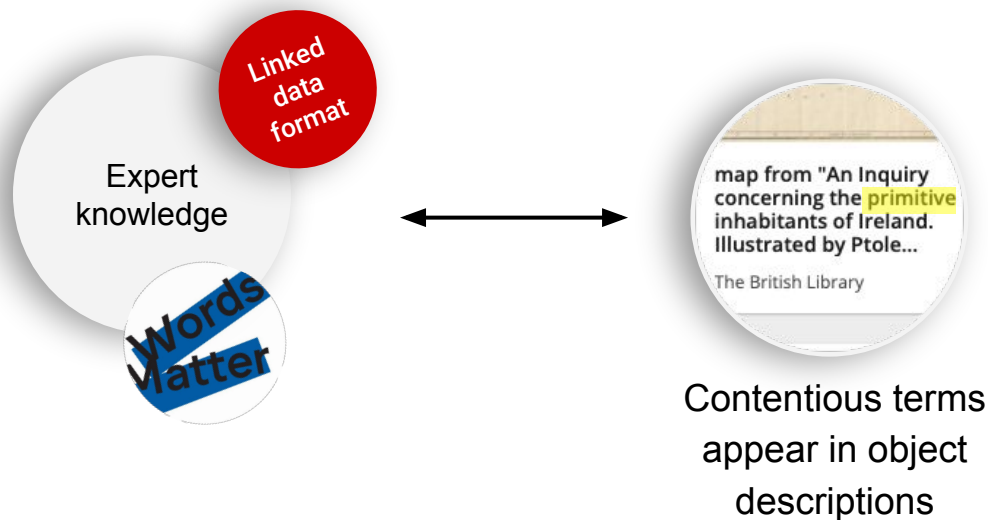
How to formalise contentiousness for knowledge representation and reasoning?

Developing an ontology and a controlled thesaurus of contentious terms based on "Words Matter"

Surveying the Sematic Web: how are contentious terms used?

Collecting more empirical evidence to identify problematic aspects of contentious terms usage

Making contentiousness machine readable



Modelling relationships
between contentious terms,
their contexts, suggestions, and
alternatives

From expert knowledge to linked data

GLOSSARY OF TERMS

Primitive

HISTORY, USE & POSSIBLE SENSITIVITIES

Primitive derives from the Latin word *primitivus*, meaning the first-born or first of its kind. In European thought, it became synonymous with the racialized and temporal "Other" and was applied to cultures that were imagined as existing in a distant past (see Marijke Kunst and Simone Vermaat in this publication) that lacked qualities that were seen as European, specifically progress and rationality.

The term is before still used today to denote someone or something as simple and uncivilized.

"Primitivism" is a movement in art that connotes the beauty, natural qualities, or romance of a simpler way of life, in contrast to a more advanced, industrialized Europe.

Both the term "primitive", like the artistic movement, have received significant criticism from diverse quarters including academics.

SUGGESTIONS

- The term is not recommended for use.
- The term can be used in a descriptive or historical context, in which case we suggest the use of quotation marks.

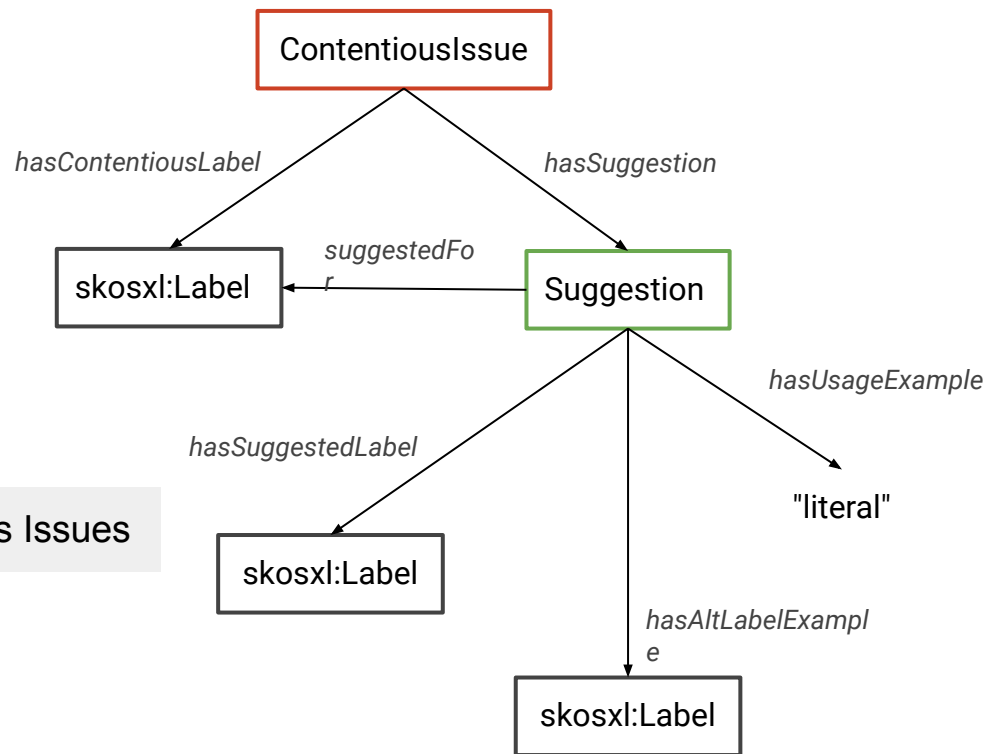
For example: *There was an artistic movement called 'primitivism'.*

132



111 Contentious Issues

296 Labels



From expert knowledge to linked data

GLOSSARY OF TERMS

Primitive

HISTORY, USE & POSSIBLE SENSITIVITIES

Primitive derives from the Latin word *primitivus*, meaning the first-born or first of its kind. In European thought, it became synonymous with the racialized and temporal “Other” and was applied to cultures that were imagined as existing in a distant past (see Marijke Kunst and Simone Vermaat in this publication) that lacked qualities that were seen as European, specifically progress and rationality.

The term is before still used today to denote someone or something as simple and uncivilized.

“Primitivism” is a movement in art that connotes the beauty, natural qualities, or romance of a simpler way of life, in contrast to a more advanced, industrialized Europe.


Both the term “primitive”, like the artistic movement, have received significant criticism from diverse quarters including academics.

SUGGESTIONS

- The term is not recommended for use.
- The term can be used in a descriptive or historical context, in which case we suggest the use of quotation marks.

For example: *There was an artistic movement called ‘primitivism’.*

132



```
wsm:132
a culco:ContentiousIssue, bibo:BookSection ;
dcterms:title "Primitive"@en ;
dcterms:description "Primitive derives from the Latin word primitivus, meaning the
first-born or first of its kind. In European thought, it became synonymous with the
racialized and temporal “Other” and was applied to cultures that were imagined as
existing in a distant past (see Marijke Kunst and Simone Vermaat in this
publication) that lacked qualities that were seen as European, specifically
progress and rationality. The term is before still used today to denote someone or
something as simple and uncivilized. “Primitivism” is a movement in art that
connotes the beauty, natural qualities, or romance of a simpler way of life, in
contrast to a more advanced, industrialized Europe. Both the term “primitive”, like
the artistic movement, have received significant criticism from diverse quarters
including academics."@en ;
culco:hasContentiousLabel
wsm:l_90 ,
wsm:l_91 ;
culco:hasSuggestion
wsm:2_s,
wsm:132_s_02 .

wsm:132_s_02
a culco:Suggestion ;
rdf:value "The term can be used in a descriptive or historical context, in which case
we suggest the use of quotation marks. For example: There was an artistic movement
called ‘primitivism’."@en ;
culco:suggestedFor
wsm:l_90,
```

Turtle syntax

SPARQL: get suggested terms for contentious

```
SELECT ?ContentiousLabel ?SuggestedLabel
WHERE {
  ?Suggestion culco:suggestedFor / skosxl:literalForm ?ContentiousLabel ;
              culco:hasSuggestedLabel / skosxl:literalForm ?SuggestedLabel .
}
```

Slaaf: mensen in slavernij
Caucasian: White
Jappenkamp: Japanese concentration camp
Jappenkampen: Japanese concentration camp
Colored: Black and Non-Black person of color
travestiet: crossdresser
Trans: crossdresser
Oriental: Asian
Blank: Wit

...



Available
on TriplyDB



Computational approaches to contentious terms

How to collect training or "ground-truth" data?

ConConCor: using crowdsourcing to construct a corpus of contentious terms in context

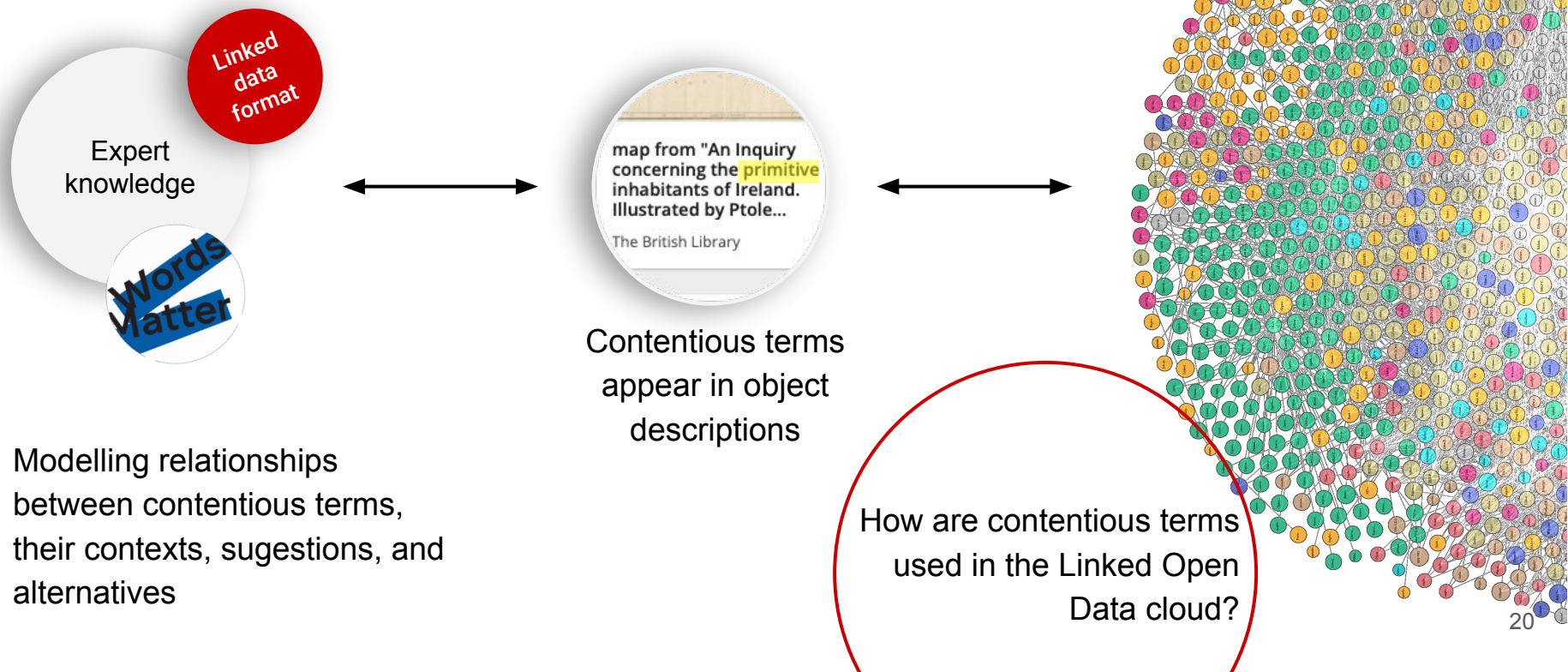
How to formalise contentiousness for knowledge representation and reasoning?

Developing an ontology and a controlled thesaurus of contentious terms based on "Words Matter"

Surveying the Sematic Web: how are contentious terms used?

Collecting more empirical evidence to identify problematic aspects of contentious terms usage

Making contentiousness machine readable



Contentious terms in Wikidata

primitive tribal community (Q13165863)

socio-economic unit of primitive society

primitive community, I primitive communal system

▼ In more languages

Configure

Language	Label	Description	Also known as
English	primitive tribal community	socio-economic unit of primitive society	primitive community, primitive communal system
Dutch	No label defined	No description defined	
German	No label defined	No description defined	
French	communauté tribale primitive	unité socio-économique d'une société primitive	

No
information
about
contentious
-ness

Used as
training
data

Wikipedia article about the term's usage

Use of the term "primitive" [\[edit \]](#)

"Primitive" in recent anthropological and social studies has begun to fall out of use due to racial stereotypes surrounding the ideas of what is primitive.^[112] Such a move has been supported by indigenous peoples who have faced racial stereotyping and violence due to being viewed as "primitive".^{[113][114]} Due to this, the term "primitive communism" may be replaced by terms such as [Pre-Marxist communism](#).^[115]

Alain Testart and others have said that anthropologists should be careful when using research on hunter-gatherer societies to determine the structure of societies in the [paleolithic](#), where viewing hunter-gatherer communities as "the most ancient of so-called primitive societies" is likely due to assumptions and perceptions and does not reflect the progress and development that such societies have undergone in the past 10,000 years.^[116]

There have been [Marxist historians](#) criticised for their comments on the "primitivism" and "barbarism" of societies prior to their contact with [European empires](#), such as the comments of [Endre Sík](#). Such views on "primitivism" and "barbarism" are also prevalent in the works of their non-Marxist contemporaries.^{[117][60][118]} Marxist anthropologists have criticised and denounced Soviet anthropologists and historians for declaring indigenous communities they were studying for primitive communism as "degenerate".^[44]

Implicit
information

Contentious terms in Wikidata

primitive data type (Q1188648)

basic data type which can be instantiated and used with the predefined operations of a programming language, or assembled as a building block together with other instances of any type to create more complex composite data types [edit](#)

primitive type | basic type | built-in type | machine type | type of the system | native type

▼ In more languages

Configure

Language	Label	Description	Also known as
English	primitive data type	basic data type which can be instantiated and used with the predefined operations of a programming language, or assembled as a building block together with other instances of any type to create more complex composite data types	primitive type basic type built-in type machine type type of the system native type
Dutch	primitief type	No description defined	primitief gegevenstype primitief datatype

Main challenge: word sense disambiguation

1. Polysemy and homonymy

The term "primitive" has 7 definitions in Princeton WordNet

2. Lack of linguistic information about the literal values in LOD

3. Automatic sense disambiguation of literals in Knowledge Graphs?

Nouns

(n) primitive - a word serving as the basis for inflected or derived forms; "pick" is the primitive from which "picket" is derived"

MORE ►

(n) primitive - a mathematical expression from which another expression is derived

MORE ►

(n) primitive person, primitive - a person who belongs to an early stage of civilization

MORE ►

Adjectives

(s) primitive - used of preliterate or tribal or nonindustrial societies; "primitive societies"

MORE ►

(s) archaic, primitive - little evolved from or characteristic of an earlier ancestral type; "archaic forms of life"; "primitive mammals"; "the okapi is a short-necked primitive cousin of the giraffe"

MORE ►

(s) crude, primitive, rude - belonging to an early stage of technical development; characterized by simplicity and (often) crudeness; "the crude weapons and rude agricultural implements of early man"; "primitive movies of the 1890s"; "primitive living conditions in the Appalachian mountains"

MORE ►

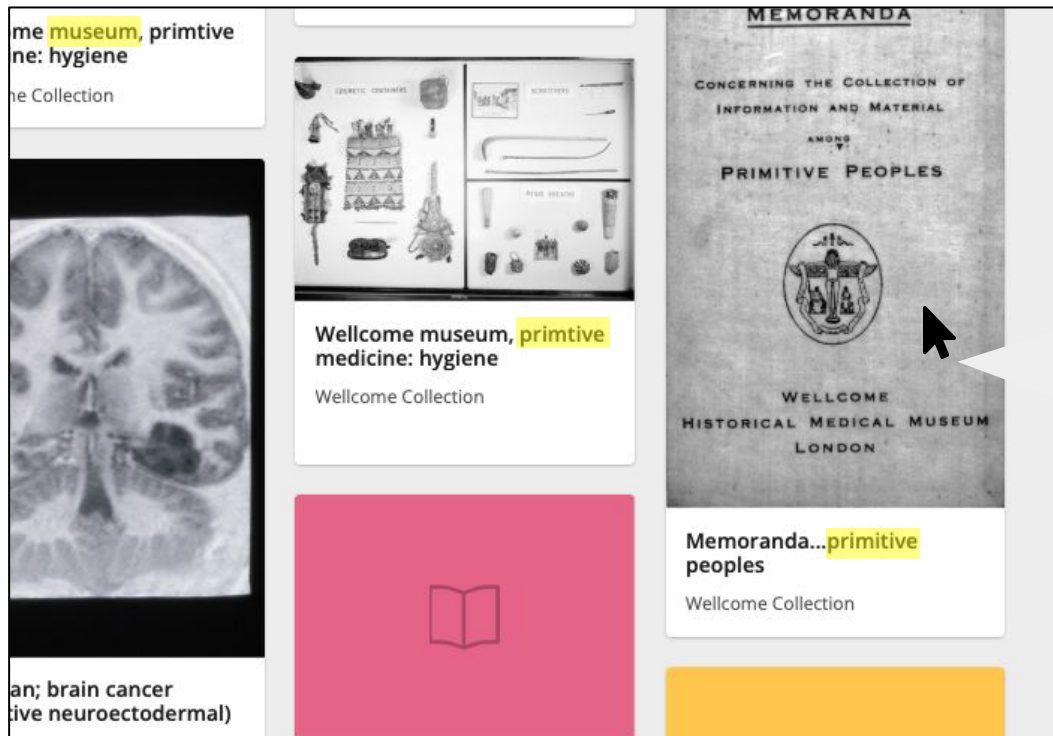
(s) primitive, naive - of or created by one without formal training; simple or naive in style; "primitive art such as that by Grandma Moses is often colorful and striking"

MORE ►

Screenshot. Princeton WordNet 3.1. Synsets of the word "primitive":

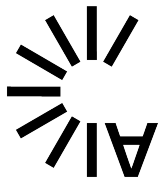
<http://wordnet-rdf.princeton.edu/lemma/primitive>

Towards the Cultural AI mission: bringing more contexts and connections to cultural heritage data



See other resources about this term:

[Words Matter](#), [Wikidata](#), [Getty AAT](#), [WordNet](#), [more](#)



Cultural AI
a lab for
culturally
valued AI



Centrum Wiskunde & Informatica



How can we develop and use computational approaches that capture, process, and represent socio-cultural sensitivities to make digital cultural heritage more inclusive?



AI for Culture



Culture for AI