LDC Catalog - Related Works Controlled Vocabulary Scheme v1.0

Terms

Resource to Resource

isSameAs

Definition

Resource A is the new/alternate name for A, while the content is identical.

Usage

Should be only used where content of resource is an exact match, such as when corpora are cross-listed with another organization. For similar, but not exact matches, other relation terms should be used.

Examples

Chinese-English Parallel Sentences Extracted from Patents (LDC2016T22) isSameAs Chinese-English Parallel Sentences Extracted from Patents (Deprecated) (LDC2016T22D). These two versions of the same catalog entry exist as an entry had to be renumbered to fix a licensing issue.

2006 CoNLL Shared Task - Ten Languages (LDC2015T11) isSameAs 2006 CoNLL Shared Task - Ten Languages (http://catalog.elra.info/en-us/repository/browse/ELRA-W0086/)

Implies

isSameAs

isSimilarWith

Definition

Resource A is similar to B in regards to creation specifications, purpose, source material, etc… or is part of a series.

Usage

This relation encapsulates resources created with similar specifications, purpose or source material. Resources may be inspired by each other or annotated at the same level, but with different tools. This also applies to corpora created as part of a series when isPartOf or isPartWith does not
apply. All resources of the same project can at least be related via
\textit{isSimilarWith}. However, with large projects, this can mean adding so many
relations as to reduce the usefulness. In cases such as these, a better use is to
relate project resources only where they are similar in regards to a second
aspect (source data, language, task, etc…) in addition to the project. Best
use is to use more specific relations where applicable.

\textbf{Examples}
BOLT Arabic Discussion Forums (LDC2018T10) \textit{isSimilarWith} BOLT Chinese
Discussion Forums (LDC2016T05).

\textbf{Implies}
\textit{isSimilarWith}

\textit{relatesTo}

\textbf{Definition}
Resource A relates to B in some broad general manner.

\textbf{Usage}
This relation should be used when no more specific relation is available, but
a definite relationship exists. For example, a corpus being inspired by
another but not with the same specific specifications, purpose, source
material, etc…

\textbf{Examples}
Phrase Detectives Corpus Version 2 (LDC2019T10) \textit{relatesTo} The ARRAU
Corpus of Anaphoric Information (LDC2013T22)

\textbf{Implies}
\textit{relatesTo}

\textbf{Resource to Resource (same type)}

\textit{isContinuationOf}

\textbf{Definition}
Resource A continues the work of resource B.

\textbf{Usage}
Rather than a new complete version of a resource which would use
\textit{isVersionOf}, \textit{isContinuationOf} describes a resource that continues work
along a specific line at a later date. \textit{isVersionOf} may be more appropriate for
cases where the corpus can operate as a fully formed corpus. Use this term in
cases of ongoing task evaluations where the data changes so dramatically
\textit{isVersionOf} doesn’t apply, but the intended task of the corpus remains
essentially the same. Similarly, this may be used to relate evaluation data to
development data or test data to training.
**Examples**
Message Understanding Conference (MUC) 6 (LDC2003T13) *isContinuationOf*
Message Understanding Conference (MUC) 6 Additional News Text (LDC96T10)

**Implies**
*hasContinuation*

**hasContinuation**

**Definition**
Resource A was continued by resource B.

**Usage**
See *isContinuationOf* for usage notes.

**Examples**
Message Understanding Conference (MUC) 6 Additional News Text (LDC96T10) *hasContinuation* Message Understanding Conference (MUC) 6 (LDC2003T13)

**Implies**
*isContinuationOf*

**isVersionOf**

**Definition**
Resource A is an extension in size, corrections of content, etc… of resource B.

**Usage**
Version relations should be used when a new resource builds on a previous one, and both are complete works in and of themselves. In cases where the differences are merely changes of format due to processing, *isOutcomeOf* may be more appropriate. Generally, in LDC usage, “editions” are covered by version relations. When the corpus developer has indicated versioning in the title, this is often a strong indicator for use of *isVersionOf* as opposed to *isContinuationOf* or *isOutcomeOf*.

In cases where multiple versions exist, versions should be related to each other as follows: V3 *isVersionOf* V2 & V1, and V2 *isVersionOf* V1. *hasVersion* relates in the opposite direction, e.g. V1 *hasVersion* V2 & V3.

**Examples**

**Implies**
*hasVersion*
hasVersion
Definition
Resource A was extended in size, corrections of content, etc… by resource B.
Usage
See hasVersion for usage notes.
Examples
Implies
isVersionOf

replaces
Definition
Resource A replaces or supersedes resource B.
Usage
Like the version relations, replaces also indicates a subsequent extension of a resource. However, unlike isVersionOf, the previous resource is no longer intended to be used.
Examples
Switchboard-1 Release 2 (LDC97S62) replaces Switchboard-1 (LDC93S7)
Implies
isReplacedBy

isReplacedBy
Definition
Resource A was replaced or supersed by resource B.
Usage
See replaces usage notes.
Examples
Switchboard-1 (LDC93S7) isReplacedBy Switchboard-1 Release 2 (LDC97S62)
Implies
replaces

Dataset to Dataset
isOutcomeOf
Definition
Resource A is the product/outcome of resource B.
Usage
Broadly, the best uses for outcome relations are 1) some form of processing has been done to change the format of a resource or 2) a portion of another corpus has been used and modified to produce the corpus in question, such as a translation. This modification may simply be taking a portion of a corpus as source data. When this processing is primarily annotation, isAnnotationOf may be more appropriate. isVersionOf may be more appropriate when substantive changes to a resource involve more than format, such as the addition or correction of content, but the intent of the corpus remains essentially the same.

Examples
TIMIT Acoustic-Phonetic Continuous Speech (MS-WAV version) (LDC93S1W) isOutcomeOf TIMIT Acoustic-Phonetic Continuous Speech Corpus (LDC93S1)

ASpIRE Development and Development Test Sets (LDC2017S21) isOutcomeOf Mixer 6 Speech (LDC2013S03)

RATS Speech Activity Detection (LDC2015S02) isOutcomeOf Fisher English Training Part 2, Speech (LDC2005S13)

Implies
hasOutcome

hasOutcome
Definition
Resource A has a product/outcome as resource B.

Usage
See isOutcomeOf for usage notes.

Examples
TIMIT Acoustic-Phonetic Continuous Speech Corpus (LDC93S1) hasOutcome
TIMIT Acoustic-Phonetic Continuous Speech (MS-WAV version) (LDC93S1W)

Implies
isOutcomeOf

isAnnotationOf
Definition
Resource A is annotation of resource B.

Usage
Annotation relations are best used when a resource applies descriptive or analytic notations to another resource, such as transcription. This second resource may or may not be contained within the annotation resource.
Examples
CALLHOME Mandarin Chinese Transcripts (LDC96T16) isAnnotationOf CALLHOME Mandarin Chinese Speech (LDC96S34)

TAC KBP English Regular Slot Filling - Comprehensive Training and Evaluation Data 2009-2014 (LDC2018T22) isAnnotationOf TAC KBP Comprehensive English Source Corpora 2009-2014 (LDC2018T03)

Implies
hasAnnotation

hasAnnotation
Definition
Resource A has annotation in resource B.

Usage
See isAnnotationOf for usage notes.

Examples
Mandarin Chinese Speech (LDC96S34) hasAnnotation CALLHOME Mandarin Chinese Transcripts (LDC96T16)

TAC KBP Comprehensive English Source Corpora 2009-2014 (LDC2018T03) hasAnnotation TAC KBP English Regular Slot Filling - Comprehensive Training and Evaluation Data 2009-2014 (LDC2018T22)

Implies
isAnnotationOf

isPartOf
Definition
Resource A is part of resource B.

Usage
isPartOf and hasPart are “parent/child” relationships--hierarchical in nature. These relations can express both one-to-one and one-to-many types of relationships. This covers the case where one resource is a whole or partial subset of another. isPartWith relates children of a common parent together. Best practice is to use this where the combination of parts creates a cohesive resource, which may or may not exist in the combined form. When there is an equally applicable relation such as isAnnotationOf or isOutcomeOf, it is advised to use those instead.

Examples
UN Parallel Text (English) (LDC94TB-1) isPartOf UN Parallel Text (Complete) (LDC94T4A)

Implies
hasPart
hasPart
Definition
Resource A has part as resource B.
Usage
See isPartOf usage notes.
Examples
UN Parallel Text (Complete) (LDC94T4A) hasPart UN Parallel Text (English) (LDC94TB-1)
Implies
isPartOf

isPartWith
Definition
Resource A and B are both parts of a third resource.
Usage
See isPartOf for usage notes.
Examples
UN Parallel Text (English) (LDC94TB-1) isPartWith UN Parallel Text (French) (LDC94TB-2)
Implies
isPartWith

Dataset to Tool
isCreatedBy
Definition
Resource A was created by tool B.
Usage
This term is used when a dataset has been created in part or full by a tool. Some examples might be a web crawler, OCR tool, term extractor, etc...
Examples
GALE Phase 4 Arabic Broadcast News Transcripts (LDC2018T14) isCreatedBy XTrans (https://www.ldc.upenn.edu/language-resources/tools/xtrans)
Implies
creates

creates
Definition
Tool A creates resource B.
Usages
See isCreatedBy for usage notes.

Examples
XTrans (https://www.ldc.upenn.edu/language-resources/tools/xtrans) creates GALE Phase 4 Arabic Broadcast News Transcripts (LDC2018T14)

Implies
isCreatedBy

isProcessedBy
Definition
Resource A was processed by tool B.
Usage
Use for any corpus that has been processed by a tool. These may include tools that annotate, analyze, edit or validate the corpus.

Examples
The ARRAU Corpus of Anaphoric Information (LDC2013T22) isProcessedBy The MMAX2 Annotation Tool (http://mmax2.net/index.html)

Implies
isProcessedBy

processes
Definition
Tool A processes resource B.
Usage
See isProcessedBy for usage notes.

Examples
The MMAX2 Annotation Tool (http://mmax2.net/index.html) processes The ARRAU Corpus of Anaphoric Information (LDC2013T22)

Implies
isProcessedBy

isManagedBy
Definition
Resource A is managed by tool B.
Usage
Use for any corpus managed by a tool. These may include tools that access, archive, display or query the corpus.

Examples
Concretely Annotated English Gigaword (LDC2018T20) isManagedBy Concrete (https://hltcoe.github.io/concrete/)
Implies

manages

Definition

Tool A manages resource B.

Usage

See isManagedBy for usage notes.

Examples

Concrete (https://hitcoe.github.io/concrete/) manages Concretely Annotated English Gigaword (LDC2018T20)

Implies

isManagedBy

Tool to Resource

requires

Definition

Tool A requires resource B.

Usage

Use in the case where a tool requires a resource to function. This may be grammar for a parser, a list of stop works, etc... if the resource is a corpus. Use also for a dependent piece of software if both A and B are tools.

Examples

LDC Word Aligner (https://www.ldc.upenn.edu/language-resources/tools/ldc-word-aligner) requires Python 2.6 (https://www.python.org/download/releases/2.6/)

Implies

isRequiredBy

isRequiredBy

Definition

Resource A is required by tool B.

Usage

See requires for usage notes.

Examples

Python 2.6 (https://www.python.org/download/releases/2.6/) isRequiredBy LDC Word Aligner (https://www.ldc.upenn.edu/language-resources/tools/ldc-word-aligner)

Implies

Requires