RDN-Boost API

The RDN-Boost API provides methods for training a boosted RDN model on a corpus of documents and for returning predicates from a test corpus.

|  |  |
| --- | --- |
| **API Method Summary** | |
| String | [**extractInformationFromCorpus**](#extractInformationFromCorpus(java.io.Fil)(Collection<String> relationsToExtract, Collection<File> testCorpus, double minimumProbabilityToBeReported)            Applies a learned model to new corpus files and returns annotated predicates of interest, one per line. |
| File | [**getPathnameForScratchSpace**](#getPathnameForScratchSpace())()            Returns the location used for large scratch files. If this is not set with setPathnameForScratchSpace, it will default to System.getProperty("java.io.tmpdir"). |
| boolean | [**setPathnameForScratchSpace**](#setPathnameForScratchSpace(java.io.File))(File pathname)            Allows the caller to set the location used for large scratch files. |
| File | [**trainOnCorpus**](#trainOnCorpus(java.util.ArrayList))(Collection <File> corpusFiles, Collection <File> annotatedFiles)            Trains on data given in annotated files and creates an RDN model.  This method is for internal use to University of Wisconsin only. |

**The following are predicates returned in the result file by extractInformationFromCorpus. (They are not methods.)** These predicates may be returned as rules, or in an equivalent format such as RDF XML. (Format to be determined.)

|  |
| --- |
| **Entity Predicate Summary** |
| [**count\_fromWisc**](#count_fromWisc)(String article, String extractedString, int beginChar, int endChar, double probability).            Represents an extracted numeric value, like a score of one team. |
| [**Date\_fromWisc**](#Date_fromWisc)(String article, String extractedString, int beginChar, int endChar, double probability).           Represents an extracted date. |
| [**NFLGame\_fromWisc**](#NFLGame_fromWisc)(String article, String extractedString, int beginChar, int endChar, double probability).            Represents a particular NFL Game extracted from the corpus. |
| [**NFLTeam\_fromWisc**](#NFLTeam_fromWisc)(String article, String extractedString, int beginChar, int endChar, double probability).           Represents a particular NFL Team extracted from the corpus. |

|  |
| --- |
| **Relation Predicate Summary** |
| [**awayTeamInGame\_fromWisc**](#awayTeamInGame_fromWisc)(String nflGameArticle, String nflGameString, int nflGameBeginChar, int nflGameEndChar, String teamArticle, String teamString, int teamBeginChar, int teamEndChar, double probability).            Represents the extracted away team for a given NFL game. |
| [**gameDate\_fromWisc**](#gameDate_fromWisc)(String dateArticle, String dateString, int dateBeginChar, int dateEndChar, String nflGameArticle, String nflGameString, int nflGameBeginChar, int nflGameEndChar, double probability).            Represents the extracted date for a particular game. |
| [**gameLoser\_fromWisc**](#gameLoser_fromWisc)(String nflGameArticle, String nflGameString, int nflGameBeginChar, int nflGameEndChar, String teamArticle, String teamString, int teamBeginChar, int teamEndChar, double probability).            Represents the extracted losing NFL team for a particular NFL game. |
| [**gameWinner\_fromWisc**](#gameWinner_fromWisc)(String nflGameArticle, String nflGameString, int nflGameBeginChar, int nflGameEndChar, String teamArticle, String teamString, int teamBeginChar, int teamEndChar, double probability).            Represents the extracted winning NFL team for a particular NFL game. |
| [**homeTeamInGame\_fromWisc**](#homeTeamInGame_fromWisc)(String nflGameArticle, String nflGameString, int nflGameBeginChar, int nflGameEndChar, String teamArticle, String teamString, int teamBeginChar, int teamEndChar, double probability).            Represents the extracted home team for a given NFL game. |

|  |
| --- |
| **API Method Detail** |

**extractInformationFromCorpus**

File **extractInformationFromCorpus**(Collection<String> relationsToExtract,

Collection<File>  testCorpus,

double minimumProbabilityToBeReported)

Applies a learned model to new corpus files and returns annotated predicates of interest, one per line.

**Parameters:**

relationsToExtract - Collection enumerating the relations to extract (i.e. Date\_fromWisc, awayTeamInGame\_fromWisc)

testCorpus - list of article files used to test the model

minimumProbability - minimum probability threshold of predicates to return

**Returns:**

String containing the predicates, one per line (that is, separated by newlines.)

**getPathnameForScratchSpace**

File **getPathnameForScratchSpace**()

Returns the location used for large scratch files.

If the pathname is not set with setPathnameForScratchSpace, it will default to System.getProperty("java.io.tmpdir"); for our scratch directory if unset

**Returns:**

Path to location of scratch files

**setPathnameForScratchSpace**

boolean **setPathnameForScratchSpace**(File pathname)

Allows the caller to set the location used for large scratch files .

**Parameters:**

pathname - path to location for scratch files

**Returns:**

true if successful, false otherwise

**trainOnCorpus**

File **trainOnCorpus**(Collection<File> corpusFiles,

Collection<File> annotatedFiles)

Trains on data given in annotated files and creates an RDN model.

This method is for internal use to University of Wisconsin only.

**Parameters:**

corpusFiles - Text article files

annotatedFiles - Files in gui\_xml or RDF\_xml format with annotations

**Returns:**

File object representing the file containing the model

|  |
| --- |
| **Entity Predicate Summary** |

**count\_fromWisc**

**count\_fromWisc**(String article,

String extractedString,

int  beginChar,

int  endChar,

double probability).

Represents an extracted numeric value, like a score of one team.

**Values:**

article - Article identifier

extractedString - String extracted from article describing count

begin - Beginning char offset of extractedString in article

end - End char offset of extractedString in article

probability - Probability of the current predicate

**Date\_fromWisc**

Date**\_fromWisc**(String article,

String extractedString,

int  beginChar,

int  endChar,

double probability).

Represents an extracted particular date.

**Values:**

article - Article identifier

extractedString - String extracted from article describing the date (ex. "May 23, 2009")

begin - Beginning char offset of extractedString in article

end - End char offset of extractedString in article

probability - Probability of the current predicate

**NFLGame\_fromWisc**

**NFLGame\_fromWisc**(String article,

String extractedString,

int  beginChar,

int  endChar,

double probability).

Represents a particular NFL Game extracted from the corpus.

**Values:**

article - Article identifier

extractedString - String extracted from article describing the NFL game

begin - Beginning char offset of extractedString in article

end - End char offset of extractedString in article

probability - Probability of the current predicate

**NFLTeam\_fromWisc**

**NFLTeam\_fromWisc**(String article,

String extractedString,

int beginChar,

int endChar,

double probability).

Represents a particular NFL Team extracted from the corpus.

**Values:**

article - Article identifier

extractedString - String extracted from article describing the NFL team

beginChar - Beginning char offset of extractedString in article

endChar - End char offset of extractedString in article

probability - Probability of the current predicate

|  |
| --- |
| **Relation Predicate Summary** |

**awayTeamInGame\_fromWisc**

**awayTeamInGame\_fromWisc**(String nflGameArticle,

String nflGameString,

int  nflGameBeginChar,

int  nflGameEndChar,

String teamArticle,

String teamString,

int  teamBeginChar,

int  teamEndChar,

double probability).

Represents the extracted away team for a given NFL game. **Currently relations exist in one article, but two article identifiers are included in case future relations span multiple articles.**

**Values:**

nflGameArticle - Article identifier

nflGameString - String extracted from article describing game

nflGameBeginChar - Beginning char offset of nflGameString in article

nflGameEndChar - End char offset of nflGameString in article

teamArticle - Article identifier (may be same as nflGameArticle)

teamString - String extracted from article describing away team

teamBeginChar - Beginning char offset of teamString in article

teamEndChar - End char offset of teamString in article

probability - Probability of the current predicate

**gameDate\_fromWisc**

**gameDate\_fromWisc**(String dateArticle,

String dateString,

int  dateBeginChar,

int  dateEndChar,

String nflGameArticle,

String nflGameString,

int  nflGameBeginChar,

int  nflGameEndChar,

double probability).

Represents the extracted date for a particular game .

**Values:**

dateArticle - Article identifier

dateString - String extracted from article describing game date

dateBeginChar - Beginning char offset of dateString in article

dateEndChar - End char offset of dateString in article

nflGameArticle - Article identifier

nflGameString - String extracted from article describing game

nflGameBeginChar - Beginning char offset of nflGameString in article

nflGameEndChar - End char offset of nflGameString in article

probability - Probability of the current predicate

**gameLoser\_fromWisc**

**gameLoser\_fromWisc**(String nflGameArticle,

String nflGameString,

int  nflGameBeginChar,

int  nflGameEndChar,

String teamArticle,

String teamString,

int  teamBeginChar,

int  teamEndChar,

double probability).

Represents the extracted losing NFL team for a particular NFL game .

**Values:**

nflGameArticle - Article identifier

nflGameString - String extracted from article describing game

nflGameBeginChar - Beginning char offset of nflGameString in article

nflGameEndChar - End char offset of nflGameString in article

teamArticle - Article identifier (may be same as nflGameArticle)

teamString - Article extracted from extractedString describing losing team

teamBeginChar - Beginning char offset of teamString in article

teamEndChar - End char offset of teamString in article

probability - Probability of the current predicate

**gameWinner\_fromWisc**

**gameWinner\_fromWisc**(String nflGameArticle,

String nflGameString,

int  nflGameBeginChar,

int  nflGameEndChar,

String teamArticle,

String teamString,

int  teamBeginChar,

int  teamEndChar,

double probability).

Represents the extracted winning NFL team for a particular NFL game .

**Values:**

nflGameArticle - Article identifier

nflGameString - String extracted from article describing game

nflGameBeginChar - Beginning char offset of nflGameString in article

nflGameEndChar - End char offset of nflGameString in article

teamArticle - Article identifier (may be same as nflGameArticle)

teamString - Article extracted from extractedString describing winning team

teamBeginChar - Beginning char offset of teamString in article

teamEndChar - End char offset of teamString in article

probability - Probability of the current predicate

**homeTeamInGame\_fromWisc**

**homeTeamInGame\_fromWisc**(String nflGameArticle,

String nflGameString,

int  nflGameBeginChar,

int  nflGameEndChar,

String teamArticle,

String teamString,

int  teamBeginChar,

int  teamEndChar,

double probability).

Represents the extracted home team for a given NFL game.

**Values:**

nflGameArticle - Article identifier

nflGameString - String extracted from article describing game

nflGameBeginChar - Beginning char offset of nflGameString in article

nflGameEndChar - End char offset of nflGameString in article

teamArticle - Article identifier (may be same as nflGameArticle)

teamString - Article extracted from extractedString describing home team

teamBeginChar - Beginning char offset of teamString in article

teamEndChar - End char offset of teamString in article

probability - Probability of the current predicate