

Robotics Systems Types - Feature #2738

Add "Part of Speech Tag" to Word

09/24/2017 07:18 PM - D. Hamann

Status:	Resolved	Start date:	09/24/2017
Priority:	Normal	Due date:	
Assignee:	D. Hamann	% Done:	100%
Category:	Type Proposal	Estimated time:	0.00 hour
Target version:	rsb-0.17		
Description			
branched from 0.15 and like to have it added to 0.15			
Related issues:			
Related to Robotics Systems Types - Feature # 2737: Add "Corenlp Dependency T...		Feedback	09/14/2017

Associated revisions

Revision ebd6d168 - 09/29/2017 12:51 PM - D. Hamann

Added field part-of-speech-tag to Word in proto/stable/rst/dialog/SpeechHypothesis.proto

fixes #2738

Signed-off-by: Jan Moringen <jmoringe@techfak.uni-bielefeld.de>

Revision 6f475928 - 09/29/2017 12:52 PM - D. Hamann

Backport: Added field part_of_speech_tag to Word in proto/stable/rst/dialog/SpeechHypothesis.proto

refs #2738

Signed-off-by: Jan Moringen <jmoringe@techfak.uni-bielefeld.de>

(cherry picked from commit ebd6d16853ff239e27fca161f7e46fd3a2bc6530)

Revision 5e87d699 - 09/29/2017 12:52 PM - D. Hamann

Added field part_of_speech_tag to Word in proto/stable/rst/dialog/SpeechHypothesis.proto

fixes #2738

Signed-off-by: Jan Moringen <jmoringe@techfak.uni-bielefeld.de>

Revision 8bee035b - 09/29/2017 01:20 PM - D. Hamann

Backport: Added field part_of_speech_tag to Word in proto/stable/rst/dialog/SpeechHypothesis.proto

refs #2738

Signed-off-by: Jan Moringen <jmoringe@techfak.uni-bielefeld.de>

History

#1 - 09/28/2017 11:07 AM - J. Moringen

- Status changed from New to Feedback
- Assignee set to D. Hamann

There are two issues with this:

1. Regardign the tagging scheme itself:
 - What is the canonical name and documentation for this tagging scheme?
 - The documentation under <http://www.coli.uni-saarland.de/projects/sfb378/negra-corpus/> suggests that the tagging scheme is called Stuttgart-Tübingen-Tagset (STTS). Is this correct?
 - If so, what is the reason for choosing this particular scheme?
 - To which languages is this scheme applicable?
2. The individual enum items were not documented.

Based on the above issues, I made the following version of the enum and field:

```
/**
 * Part-of-speech tags using a modified version of the
 * Stuttgart-Tübingen-Tagset (STTS).
 *
 * Differences w.r.t. STTS are:
 *
 * * ``KOMM`` instead of ``$,``
 * * ``END`` instead of ``$.``
 * * ``IPNCT`` instead of ``$(``
 *
 * @see http://www.coli.uni-saarland.de/projects/sfb378/negra-corpus/stts.asc
 * "Description of the STTS (in German)"
 */
enum PosTag {

    /**
     * Attributives Adjektiv.
     */
    ADJA = 1;

    /**
     * Adverbiales oder prädikatives Adjektiv.
     */
    ADJD = 2;

    /**
     * Adverb.
     */
    ADV = 3;

    /**
     * Präposition; Zirkumposition links.
```

*/

APPR = 4;

/**

* Präposition mit Artikel.

*/

APPRART = 5;

/**

* Postposition.

*/

APPO = 6;

/**

* Zirkumposition rechts.

*/

APZR = 7;

/**

* Bestimmter oder unbestimmter Artikel.

*/

ART = 8;

/**

* Kardinalzahl.

*/

CARD = 9;

/**

* Fremdsprachliches Material.

*/

FM = 10;

/**

* Interjektion.

*/

ITJ = 11;

/**

* Ordinalzahl.

*/

ORD = 12;

/**

* Unterordnende Konjunktion.

*/

KOUI = 13;

/**

* Unterordnende Konjunktion.

*/

KOUS = 14;

/**

* Nebenordnende Konjunktion.

*/

KON = 15;

/**

* Vergleichskonjunktion.

*/

KOKOM = 16;

/**

* Normales Nomen.

*/

NN = 17;

/**

* Eigennamen.

*/

NE = 18;

/**

* Substituierendes Demonstrativpronomen.

*/

PDS = 19;

/**

* Attribuierendes Demonstrativpronomen.

*/

PDAT = 20;

/**

* Substituierendes Indefinitpronomen.

*/

PIS = 21;

/**

* Attribuierendes Indefinitpronomen ohne Determiner.

*/

PIAT = 22;

/**

* Attribuierendes Indefinitpronomen mit Determiner.

*/

PIDAT = 23;

/**

* Irreflexives Personalpronomen.

*/

PPER = 24;

/**

* Substituierendes Possessivpronomen.

*/

PPOSS = 25;

/**

* Attribuierendes Possessivpronomen.

*/

PPOSAT = 26;

/**

* Substituierendes Relativpronomen.

*/

PRELS = 27;

/**

* Attribuierendes Relativpronomen.

*/

PRELAT = 28;

/**

* Reflexives Personalpronomen.

*/

PRF = 29;

/**

* Substituierendes Interrogativpronomen.

*/

PWS = 30;

/**

* Attribuierendes Interrogativpronomen.

*/

PWAT = 31;

/**

* Adverbiales Interrogativ- oder Relativpronomen.

*/

PWAV = 32;

/**

* Pronominaladverb.

*/

PAV = 33;

/**

* "zu" vor Infinitiv.

*/

PTKZU = 34;

/**

* Negationspartikel.

*/

PTKNEG = 35;

/**

* Abgetrennter Verbzusatz.

*/

PTKVZ = 36;

/**

* Antwortpartikel.

*/

PTKANT = 37;

/**

* Partikel bei Adjektiv oder Adverb.

*/

PTKA = 38;

/**

* SGML Markup.

*/

SGML = 39;

/**

* Buchstabierfolge.

*/

SPELL = 40;

/**

* Kompositions-Erstglied.

*/

TRUNC = 41;

/**

* Finites Verb, voll.

*/

VVFIN = 42;

/**

* Imperativ, voll.

*/

VVIMP = 43;

/**

* Infinitiv, voll.

*/

VVINFIN = 44;

/**

* Infinitiv mit "zu", voll.

*/

VVIZU = 45;

/**

* Partizip Perfekt, voll.

*/

VVPP = 46;

/**

* Finites Verb, aux.

*/

VAFIN = 47;

/**

* Imperativ, aux.

*/

VAIMP = 48;

/**

* Infinitiv, aux.

*/

VAINF = 49;

/**

* Partizip Perfekt, aux.

*/

VAPP = 50;

/**

* Finites Verb, modal.

*/

VMFIN = 51;

/**

* Infinitiv, modal.

*/

VMINF = 52;

/**

* Partizip Perfekt, modal.

*/

VMPP = 53;

/**

* Nichtwort, Sonderzeichen enthaltend.

*/

XY = 54;

/**

* Komma

*/

KOMM = 55;

/**

* Satzbeendende Interpunktion.

*/

END = 56;

/**

* Sonstige Satzzeichen; satzintern.

```
*/
IPNCT = 57;

}

/**
 * Part-of-speech tag for this word.
 */
optional PosTag part_of_speech_tag = 3;
```

Apart from the questions raised in 1., would this OK?

#2 - 09/28/2017 05:48 PM - D. Hamann

J. Moringen wrote:

There are two issues with this:

1. Regardign the tagging scheme itself:

- *What is the canonical name and documentation for this tagging scheme?*
- *The documentation under <http://www.coli.uni-saarland.de/projects/sfb378/negra-corpus/> suggests that the tagging scheme is called Stuttgart-Tübingen-Tagset (STTS). Is this correct?*
- *If so, what is the reason for choosing this particular scheme?*
- *To which languages is this scheme applicable?*

Yes, this tagging scheme is called Stuttgart-Tübingen-Tagset (STTS). It is only applicable for german language.

This is the default tag-set and it has the best performance from the the german data-model for the toolkit from which we recieve the tags.

Info about the tagger-models: <https://github.com/stanfordnlp/CoreNLP/blob/master/doc/tagger/README-Models.txt>

- 1. The individual enum items were not documented.*

Based on the above issues, I made the following version of the enum and field:

[...]

Apart from the questions raised in 1., would this OK?

Yes this is OK.

#3 - 09/29/2017 01:00 PM - D. Hamann

- *Status changed from Feedback to Resolved*
- *% Done changed from 0 to 100*

Applied in changeset commit:rst-protoljebd6d16853ff239e27fca161f7e46fd3a2bc6530.

#4 - 09/29/2017 01:15 PM - J. Moringen

- *Target version set to rsb-0.17*

- % Done changed from 100 to 0

#5 - 09/29/2017 01:19 PM - D. Hamann

D. Hamann wrote:

| *Applied in changeset commit:rst-protol|ebd6d16853ff239e27fca161f7e46fd3a2bc6530.*

I did not do that. How does this happen from my account!?

#6 - 09/29/2017 01:27 PM - J. Moringen

| *I did not do that. How does this happen from my account!?*

Sorry, this confuses everybody. I revised your patch, keeping the Author-Metadata of the commit intact, then pushed it. I also annotated in the commit message that the commit fixes this issue. Redmine interprets that as you closing the issue because you are the Author of the commit.

#7 - 09/29/2017 01:43 PM - J. Moringen

- % Done changed from 0 to 100

#8 - 09/29/2017 02:56 PM - J. Moringen

- *Related to Feature #2737: Add "Corenlp Dependency Tree" to SpeechHypothesis added*

Files

0001-added-part-of-speech-tag-to-word.patch	2.87 KB	09/24/2017	D. Hamann
---------------------------------------------	---------	------------	-----------