Czech Stickers

Daniel Zeman, Jan Hajič
Charles University in Prague, Institute of Formal and Applied Linguistics (ÚFAL)
Univerzita Karlova v Praze, Ústav formální a aplikované lingvistiky (ÚFAL)
Malostranské náměstí 25, Praha, CZ-11800, Czechia
{zeman|hajic}@ufal.mff.cuni.cz

Motto

By 2040, the ACL conference proceedings shall be machine translated into 20 languages from the original Chinese.

Kevin Knight[1]

1 介紹捷克標籤系統

When the development of a lexical analyzer for a given language, it is the first to identify a set of possible label, consistent with our language ideology. Každý záznam bude obsahovat tyto informace (v obecném slova smyslu), o gramatické kategorie slovní forma v otázce, která patří k morfologické úrovni jazyka popsal. Das ist etwas vage Definition des Unterschieds zwischen den Ebenen, obwohl intuitiv klar, technisch schwierig zu definieren. Il existe de nombreux cas limites. El tudod képzelni, például, hogy "része a beszéd" kategória, mint a kategória nem morfológiai, de több okból kifolyólag, hagyományosan része a morfológiai tag. Por otro lado, se podría decir que incluso las funciones sintácticas posibles (como "sujeto", "sujeto" o una "filial"), la palabra puede ser parte del día (ver la sintaxis de marcado "Por último", el texto completo del análisis es completo, la palabra – o lexema se basará en – tal función, por supuesto. Is-sistema hija membru tal-trasformazzjoni morfoloġiċi Ċek isir kif deskritt, kienu l-istess, it-tnejn kienu jiffurmaw punt ta 'referenza. Один вниз, один за другим символы на морфологическом (получить общую сумму морфологических категорий сформировать конкретное слово в контексте). Yksi niistä on nimeltään "sopimus" toinen "asentoon". Do podpisania porozumienia między morfologicznych Czech zajmuje mniej miejsca (stąd nazwa) i, co najważniejsze, jest to łatwiejsze do zapamiętania i zarządzania, a słownik osób. ब्रांड पोजिशनिंग सिस्टम, लेकिन यह भी सीधे लागू करने से पहले Tagger (6.3 अनुभाग देखें). क्योंकि वह स्तंभ तय की है (या "पोस्ट") प्रत्येक morphological समूह के लिए. Nevyužitá ("neuplatňuje sa" alebo "N. 1.") hodnota je použitie špeciálnej markér znakov (pomlčka, '-'). GPS är för närvarande den mest populära märken, och vårt arbete, som omfattar alla konformation Tjeckien. Ông (và đánh bại), hai thương hiệu hàng đầu trên thế giới, nhưng để biết được nhân viên. Gerçek kökenli iki önceki sürümleri hesaplama sözleşme sistemine açıklanan. Ако е необходимо, можете да конвертирате данните. Există întâlniri cu Cehă. Кращий спосіб виразити вид та тип карти відправити нова інформація може бути отримана. Svona, nú á öllum reikningum og önnur verkefni sem tengjast landafræði, sagði í Tékklandi. Éigeandála i go leor réimsí a phróiseáil teanga nádúrtha, agus ar an múnla (agus ar a thionchar ar an chuma ar an ainm), go soiléir ar an todhchaí GPS p TAG (tíreolaíochta), a bhfuil foinse mhór de chineál ithreach (4,8) agus Windows imreoir (4.7), ie cuid bheag de an t-ainm a bheith. Eredu hau aurkeztu (baina seguru gora) bateragarria da.

[1] Attributed to Kevin in his bogus platform for ACL presidential election, http://www.isi.edu/natural-language/people/acl2.html

2 Translation

Morphological analysis of the evolution of the language, you first have a lot of restrictions that are compatible with our understanding of the forms of language. Each record consists of the following (included in general terms), descriptive words in the form of grammatical categories, which include morphological language. Although a few blue words, technically, it is difficult to obtain a clear difference. There are many controversial issues. For example, volume control and some of the "normal" type, you can imagine. Grammatical func-

tion, probably (though it is, “estimates”, “title”, or “TP”) and Japan (see words in the Bible, it is easy to linguistic analysis soon – or – courses, including the role of signing. Morphological changes in Bohemia and two points, including the same system configuration are described. One of them will describe a sequence of symbols to morphological days (forthcoming, a combined value of morphological categories relevant to form for a particular word in context). One is the so-called “contract”, another “location”. To sign the contract between the morphological Czech Republic takes up less space (hence the name) and, more importantly, it will be easier to remember and manage, while in the dictionary people. Brand Positioning System, but is directly applicable in the latest Tagger (see section 6.3.) since it uses a fixed column (or “position”) for each class of morphological characters. Car maintenance by a private group (or wrong) special characters (diagnosis). GPS is currently the most popular brand, and all forms, including the Czech Republic is our work. Tools for processing, even if they are familiar with the brand position for normal operation, the output of the system must create (and entrance), two systems, through their representatives. Calculating the true origin of the two previous versions of the contract system is described. If necessary, change the data. Czech Positioning Conference. The best way is to send new information that is available to represent the type of card. Thus, for future reference, and all accounts and other projects related to physical geography, as described in the Czech Republic. Suppose further that in the near future there will be several categories of brand positioning, the important task of natural language processing, appearance (visual and morphological effects of the current paradigm of the name), company name (for example, “geographical indication”), which is currently only a fraction the concept of land (4,8), style, now part of the window style (4,7), the source of information in text form, etc. This expansion will mark out the current position of the same (even though there may be out of date) upward compatibility.

3 Conclusion

- Everyday is like a line of 15 characters from the image.
- Each position (also known as “pillar”), in the string corresponds to a morphological category.
- This can also be said that any noun can deny the Czech Republic.
- This can be a form of “fraud” in respect of brand watch.
- Zájmena ve třetí osobě množného čísla se liší podle pohlaví a číslo smlouvy.
- Wake up, Lima is often associated with morphological features.
- Distribution of light is constant and close to the domestic market.

3.1 Bonus Track: Some Gender Studies

- The relationship between income and sex grammar, for example, on the net.
- Česká gramatický rod má za to, že čtyři různé hodnoty: mužský živý, neživý mužský, ženský, a kastrovat.
- Czech grammatical gender, for four different values: the dead people, women, men, and the neutral.
- Czech Genealogy, women and four types of values and neutral, respectively, for killing people.
- Czech Genealogy, four women and a neutral value, and the death of his friend.
- České a Francie pohlaví jsou myšlenka mít čtyři různé hodnoty: muž animace, bez života, muži, ženy a Neutral.
- Ve většině ostatních případů animateness mužských forem je určena také pomocí I pro neživé mužské pohlaví a M pro mužské pohlaví animovat.

Acknowledgements

მადლობა Google Translate შემოქმედებითი მიდგომა ენაზე. საკვები და ტანსაცმელი და ავტორს ამ ქაღალდის უზრუნველყოფს გრანტის MSM0021620838 ჩეხეთის განათლების სამინისტრო.

References

Jan Hajič: Disambiguation of Rich Inflection. Computational Morphology of Czech. Karolinum, Praha, Czechia, 2004, chapter 2, pages 31 – 69.

1月哈伊奇：消富拐點。計算形態捷克。Karolinum，布拉格，捷克，2004年，第2章，頁 31 - 69。

Januar Hajic: Verbraucher reichen Wendepunkt. Berechnete Form der Tschechischen Republik. Karolinum, Prag, Tschechische Republik, 2004, Kapitel 2, Seite 31-69.

Januára obhajujúce: Spotrebitelia bohaté bodu zlomu. Vypočítaná tvorí Česká republika. Karolinum, Praha, Česká republika, 2004, kapitola 2, str 31 až 69.

Jan Hajič: Rozcestníky z Rich skloňování. Výpočetní Morfologie českých. Karolinum, Praha, Česko, 2004, kapitola 2, str. 31 až 69.

Defending John: Disambiguation of Rich inflection. Computational morphology of Czech. Wiley, Prague, Czech Republic, 2004, Chapter 2, page 31 to 69

Védekezés János: Egyértelműsítő gazdag inflexiós. morfológiája kiszámítása a Cseh Köztársaságban. Wiley, Prága, Cseh Köztársaság, 2004, 2. fejezet, 31-69 oldal

The defense of John: Disambiguation rich inflection. Calculation of the morphology of the Czech Republic. Wiley, Prague, Czech Republic, 2004, Chapter 2, p. 31-69

Id-difiża ta 'John: diżambigwazzjoni inflessjoni sinjuri. Kalkolu tal-morfoloġija tar-Repubblika Ċeka. Wiley, Praga, ir-Repubblika Ċeka, 2004, Kapitolu 2, p. 31-69

Защита от Иоанна: значения богатых перегиба. Расчет морфологии Чешской Республики. М., Прага, Чешская Республика, 2004, глава 2, стр. 31-69

Protection from John: the values of the rich inflection. Calculation of the morphology of the Czech Republic. Moscow, Prague, Czech Republic, 2004, chapter 2, p. 31-69

Suoja John: arvot rikkaiden taivutuksessa. Laskeminen morfologiaa Tšekki. Moskova, Praha, Tšekki, 2004, luku 2, s. 31-69

John ochronne: wartości bogaty przegięcia. Obliczanie morfologii Republiki Czeskiej. Moskwa, Praga, Republika Czeska, 2004, Rozdział 2, str. 31-69

John protective value of the rich inflection. Calculation of the morphology of the Czech Republic. Moscow, Prague, Czech Republic, 2004, Chapter 2, p. 1931-1969

जॉन अमीर मोड़ के सुरक्षात्मक मूल्य. चेक गणराज्य के आकृति विज्ञान की गणना. मॉस्को, प्राग, चेक गणराज्य, 2004, अध्याय 2, पी. 1931-1969

John Rich turn the protective value. Drainage from the account of the Czech Republic. Moscow, Prague, Czech Republic, 2004, Chapter 2, p. 1931-1969

John Rich tur den skyddande värde. Dränering från kontot i Tjeckien. Moskva, Prag, Tjeckien, 2004, kapitel 2, s. 1931-1969

존 리치는 보호 가치를 켜십시오. 배수 계정에서 체코합니다. 모스크바, 프라하, 체코, 2004, 제 2 장, 페이지 1931에서 1969 사이의

John Rich turn the protected value. Multiple accounts are in the Czech Republic. Moscow, Prague, Czech Republic, 2004, Chapter 2, page of 1931-1969

John Rich korunan değeri açın. Birden fazla hesap Çek Cumhuriyeti bulunmaktadır. Moskova, Prag, Çek Cumhuriyeti, 2004, Bölüm 2, 1931-1969

Джон Рич превърне защитени стойност. Има много сметки на Чешката република. Москва, Прага, Чешка република, 2004 г., част 2, стр. 1931-1969

John Rich rândul său, protejate de valoare. Există mai multe conturi din Republica Cehă. Moscova, Praga, Republica Cehă, 2004, partea 2, p. 1931-1969

Джон Річ свою чергу, охоронювані цінності. Є кілька рахунків в Чеській Республіці. Москва, Прага, Чеська Республіка, 2004, частина 2, стор 1931-1969

John Rich snúa, vernda gildi. Það eru nokkrir reikningar í Tékklandi. Moscow, Prag, Tékklandi, 2004, Part 2, p. 1931-1969

John Rich turn, protect the force. There are several accounts in the Czech Republic. Moscow, Prague, Czech Republic, 2004, Part 2, p. 1931-1969

ジョンリッチターンは、力を保護します。がチェコ共和国のいくつかのアカウントがあります。モスクワ、プラハ、チェコ共和国、 2004年第2頁 1931-1969

Jonritchitan will protect the force. There are several accounts of the Czech Republic. Moscow, Prague, Czech Republic, the second page in 2004 1931-1969

Jonritchitan indarra babestu egingo ditu. Daude Txekiar Errepublikako hainbat kontu. Mosku, Praga, Txekiar Errepublika, bigarren orrian 2004 1931-1969

Jonritchitan power will protect. There are many things about the Czech Republic. Moscow, Prague, Czech Republic, the second on the 2004 1931-1969

Jonritchitan δύναμη θα προστατεύει. Υπάρχουν πολλά πράγματα για την Τσεχική Δημοκρατία. Μόσχα, Πράγα, Τσεχική Δημοκρατία, η δεύτερη για τον κωδικό 2004 1931-1969

Jonritchitan macht zal beschermen. Er zijn veel dingen over de Tsjechische Republiek. Moskou, Praag, Tsjechische Republiek, de tweede voor de code 2004 1931-1969

Jonritchitan vai poder proteger. Há muitas coisas sobre a República Checa. Moscou, Praga, República Checa, o segundo para o código 2004 1931-1969

Beidh Jonritchitan a bheith in ann a chosaint. Tá a lán rudaí faoi Poblacht na Seice. Moscó, Prág, Poblacht na Seice, an dara ceann leis cód 2004 1931-1969

Jonritchitan bodo mogli zaščititi. Obstaja veliko stvari o Češki republiki. Moskva, Praga, Češka republika, druga pa s kodo 2004 1931-1969

Jonritchitan be able to save. There are many things about the Czech Republic. Moscow, Prague, Czech Republic, the other with code 2004 1931-1969

Jonritchitan essere in grado di salvare. Ci sono molte cose circa la Repubblica ceca. Mosca, a Praga, Repubblica Ceca, l'altra con il codice 2004 1931-1969

Jonritchitan varēs ietaupīt. Ir daudzas lietas par Čehiju. Maskava, Prāga, Čehija, kas ar kodu 2004 1931-1969

Jonritchitan sẽ có thể tiết kiệm. Có rất nhiều điều về Cộng hòa Séc. Moscow, Prague, Cộng hòa Séc, trong đó với mã 2004 1931-1969

Jonritchitan će se moći spasiti. Postoje mnoge stvari o Češkoj Republici. Moskva, Prag, Češka Republika, u kojoj broj 2004 1931-1969

Jonritchitan budou moci zachránit. Existuje mnoho věcí, o Česká republika. Moskva, Praha, Česká republika, v nichž se počet 2004 1931-1969

Et al.

4 You Still Did Not Stand It Under?

For the non-MT people, here is a solution to the first two sections:

When developing a morphological analyzer for a given language, it is necessary first to define a set of possible tags, which correspond to our linguistic notion of morphology. Each tag will contain such information (in the general sense) about the grammatical categories of the word form in question, which belong to the morphological level of the language described. This is a rather vague definition as the distinction among levels is, although intuitively clear, technically difficult to define. There are many borderline cases. One can imagine, for example, that the "part of speech" category will not be considered a morphological category, but for many reasons it is traditionally made part of the morphological tag. On the other hand, one might say that even possible syntactic functions (such as "Object", "Subject" or "Adverbial") of a given word can be made part of a tag (cf. the notion of "syntactic tagging"; eventually, after the full parse of a text is finished, the word – or its underlying lexeme – will acquire such a function, of course. In the tag system developed for the Czech morphological processing described here, two equivalent tag notation systems have been developed. Either of them uses a string of symbols to denote a morphological tag (i.e., a combination of values of morphological categories relevant for a given word form in context). One of them is called "compact", the other one "positional". The compact tag system is used in the Czech morphological dictionary, since it takes less space (hence its name) and, more importantly, is easier to remember and handle when maintaining the dictionary by humans. The positional tag system, however, is directly usable in the newest tagger (see Sect. 6.3), as it uses one stable column (or "position") in a tag for each morphological category. Unused ("not applicable", or "N.A.") values are marked using a special symbol (a hyphen, `'-'`). The positional system is now the preferred notational system in all our work involving Czech morphology. The processing tools, although operating by default in the positional tag system, can be set to provide output (and accept input) in both notational systems. Both notational systems are described here, as the actual source of all the various versions of the dictionary is being maintained using the compact tag system. Conversion information is also provided where appropriate. The Positional Tag System for Czech. The Positional Tag System is a newer and preferred method of recording the morphological information associated with the usage of word forms. Therefore, it should be used in the future for all new annotation and other projects involving Czech morphology described here. We also suppose that in the near future, the positional tag system will be extended to cover more categories important to natural language processing tasks, such as aspect (which in fact does have some morphological implications, currently hidden in the paradigm naming scheme), name entity type (such as "geographical name"), currently only part of the Term field (Sect. 4.8), style features, now part of the Style field (Sect 4.7), word-forming derivation information etc. Such an extension will leave the current tag positions intact (even though possibly obsolete) for upward compatibility.