

Решения по препроцессингу текстов

(Языки: английский, португальский, испанский, индонезийский, польский, чешский, литовский).

Команда

- Анастасия Кузнецова
- Анна Лapidус
- Альфия Бабий
- Надежда Катричева

Сначала мы взялись делать вариант В задания, но походу решения задач этого варианта, мы почувствовали в себе желание обучить классификатор, поэтому в нашей выборке для классификатора оказалось 7 языков.

1 Общие замечания

- Для удобства работы с текстами необходимо приведение к нижнему регистру;
- Удаление знаков препинания (всех, так как для нашей задачи сохранение слов с дефисом/апострофом скорее факультативно);
- Удаление цифр;
- Удалить все после обозначения рубрики “Ссылки” (References), т.к. часто они ссылаются на источники на других языках;
- Удаляем аббревиатуры, так как они выходят в топ списка частотных n-грамм. Аббревиатуры не являются словами, поэтому порядок символов в них может не соответствовать естественной последовательности символов в данном языке.

2 Предобработка

Учитывая особенности наших языков, мы решили написать общую функцию предобработки для тех, где нет особенностей, которые стоит учитывать при токенизации. С помощью этой функции можно обработать *испанский, португальский, чешский*. Он подойдет также для *польского и литовского* с небольшими изменениями, описанными ниже.

2.1 Индонезийский язык

Присутствует гортанная смычка в конце некоторых слов в виде апострофов, она является значимой частью слова, но поскольку мы будем обрабатывать n-граммы, а не слова, и апостроф будет давать лишний уникальный символ для этого языка, мы удаляем его вместе с другими знаками препинания.

2.2 Литовский язык

Можно токенизировать по пробелам, но есть сложности с обозначениями года и месяца в датах: **g. 1992 m. rugsėjis 6 d.**, есть заимствования/слова из других языков: **Casa d'Arte Bragaglia**. И если заимствованные названия встречаются редко, то слово год "metai" и слово "diena" день, которые сокращаются до m. d., если их не убрать, станут самыми частотными словами (и униграммами) в литовском словаре. Слово "Ссылка" тоже нужно задать специальное для этого языка (nuorodos).

2.3 Португальский язык

Ничего, что мешало бы определению языка, не обнаружено, и мы можем пренебречь некоторыми словами, которые пишутся через дефис, однако отдельно не употребляются: например, **tê-lo, renderam-lhe** итп.

2.4 Польский язык

Хорошо токенизируется по пробелам. Сложных случаев для токенизации не было найдено, только слово "Ссылка" тоже нужно задать специальное для этого языка (przypisy).

2.5 Чешский язык

В чешском языке по большей части нет отличительных особенностей, требующих отдельных правил при токенизации. Есть возвратная частица **se** (или **si**), которую хотелось бы соединить с глаголом в один токен, т.к. без возвратной частицы глагол имеет другое значение, но **se** может встречаться не только в качестве возвратной частицы, поэтому ввести регулярное выражение для объединения **se** и глагола не получится. Проблемы при токенизации могут быть связаны с наличием статей с большим количеством слов на другом языке (например, статья про баскско-исландский пиджин на чешском языке). Однако в общей массе текстов эти слова не должны оказаться наиболее частотными, а избавиться от них в процессе предобработки не представляется возможным.

2.6 Испанский язык

Судя по всему, испанский – крайне удобный язык для предобработки: ни дефисов, ни апострофов, ни гортанных смычек или других особенностей фонетики, обозначаемых чем-то напоминающим знаки препинания, в просмотренных текстах не встретилось.

2.7 Английский язык

Удаляем **[’s]**, он оказался слишком частотным, больше ничего удалять не представляется необходимым.

Примеры текстов для языков можно найти ниже.

3 Результаты

- По матрице схожести для 7 языков можно увидеть, что наиболее похожими языками по частотным триграммам оказываются испанский и португальский (65%

общих триграмм), что вполне ожидаемо. Высокий процент совпадения триграмм у испанского и английского (48%). Славянские языки тоже оказываются похожими (34% схожих триграмм у чешского и польского). Литовский язык похож на испанский и индонезийский.

- Нам оказалось достаточно 700 текстов, совсем без ошибок получилось при классификации на триграммах.
- Очень похоже, что наша классификация оказалась такой успешной именно благодаря уникальным символам: уже при 2-граммах точность составила 0.99, при 3-граммах она была уже 1.00. Очевидно, что классификатор выделял особые символы в качестве фиш, так как более всего ошибался в определении английского языка, в котором уникальных символов нет.

4 Примеры текстов

4.1 Английский

Doc1. Ada Maddocks.txt

Ada Maddocks Ada Maddocks (22 September 1927–7 March 2007) was a British trade union official. Born in Felixstowe, Maddocks grew up in Walthamstow and was educated at Walthamstow County High School before working in a laboratory for the Co-operative Wholesale Society. She became active in the co-operative movement, then began working for the National and Local Government Officers Association in 1961. In 1976, she assumed lead responsibility for health service employees in the union, and was elected to one of two women-only seats on the General Council of the Trades Union Congress. She was awarded an OBE in 1982, and served as President of the Trades Union Congress in 1990, actively supporting the ambulance workers' strike. She retired in 1992. Ada Maddocks (22 September 1927–7 March 2007) was a British trade union official.

Born in Felixstowe, Maddocks grew up in Walthamstow and was educated at Walthamstow County High School before working in a laboratory for the Co-operative Wholesale Society. She became active in the co-operative movement, then began working for the National and Local Government Officers Association in 1961. In 1976, she assumed lead responsibility for health service employees in the union, and was elected to one of two **women-only** seats on the General Council of the Trades Union Congress. She was awarded an OBE in 1982, and served as President of the Trades Union Congress in 1990, actively supporting the ambulance workers' strike. She retired in 1992.

== References ==

Doc2. 2016 Grand Prix SAR.txt

2016 Grand Prix SAR La Princesse Lalla Meryem The 2016 Grand Prix SAR La Princesse Lalla Meryem was a women's professional tennis tournament played on clay courts. It was the 16th edition of the tournament and part of the WTA International tournaments category of the 2016 WTA Tour. It took place in Rabat, Morocco, on 25–30 April 2016. The 2016 Grand Prix SAR La Princesse Lalla Meryem was a women's professional tennis tournament played on clay courts. It was the 16th edition of the tournament and part of the WTA International tournaments category of the 2016 WTA Tour. It took place in Rabat, Morocco, on 25–30 April 2016.

== Point distribution ==

== Singles main draw entrants ==
 ==== Seeds ==== 1 Rankings as of April 18, 2016
 ==== Other entrants ==== The following players received wildcards into the singles main draw:
 Ghita Benhadi
 Anna Blinkova
 Ons Jabeur
 The following players received entry as qualifiers:
 Marina Erakovic
 Kaia Kanepi
 Aleksandra Krunić
 Sílvia Soler Espinosa
 The following player received entry as a special exempt:
 Kateryna Kozlova
 The following player entered by a protected ranking:
 Laura Robson
 The following players received entry as lucky losers:
 Richèl Hogenkamp
 Anastasia Rodionova
 Sara Sorribes Tormo
 ==== Retirements ====
 Lesia Tsurenko
 == Doubles main draw entrants ==
 ==== Seeds ====
 Rankings as of April 18, 2016
 ==== Other entrants ====
 The following pairs received wildcards into the doubles main draw:
 Ghita Benhadi / Ons Jabeur
 Salma Charif / Abir El Fahimi
 ==== Withdrawals ====
 During the tournament
 Daniela Hantuchová
 == Champions ==
 ==== Singles ====
 Timea Bacsinszky def. Marina Erakovic, 6–2, 6–1
 ==== Doubles ====
 Xenia Knoll / Aleksandra Krunić def. Tatjana Maria / Raluca Olaru 6–3, 6–0.
 == References ==
 == External links ==
 Official website

Doc3. Betrest.txt

Betrest Betrest (also read as Batyires, and Batires) was a Queen Consort of Ancient Egypt. She lived during the 1st Dynasty. Betrest (also read as Batyires, and Batires) was a Queen Consort of Ancient Egypt. She lived during the 1st Dynasty.

== Name ==

Flinders Petrie may have considered the first two glyphs as part of a title, and reads the name on the Cairo stone fragment as Tarsset. Henri Gauthier reads Tef-ti-iriset, I.E.S. Edwards und Toby Wilkinson read Bat-iry-set. Today her name is commonly read as Betrest

or Batyires. According to Silke Roth the Name Batyires means "may Bata be favorably disposed toward her". She thinks that the queen's name was connected to the ancestor-deity Bata (also read as Baty). Toby Wilkinson instead translates the name with "motherhood is her companion" and points to the possible position of the queen as a mother of a king who followed her husband, king Den or Anedjib.

== Identity ==

Betrest is said to have been the mother of Semerkhet. Her name appears in Line III on the Cairo stone fragment C1, where she bears the title Mut (meaning "mother"). The identity of her husband is disputed. Some consider King Den to have been her husband. If so, King Anedjib would have been a (half-)brother of king Semerkhet. Another theory is that Betrest was the wife of the short-ruled Anedjib.

She is possibly also identified on a stela found at Abydos. The name of the person on the stela included a ram-hieroglyph (which commonly reads "Ba") and the signs "s" and "t" are visible. If this monument belongs to Queen Betrest, then it also preserves part of a title with a Horus-falcon sign, which may be part of the She Who Sees Horus title, which is a common title for queen consorts in the Old Kingdom of Ancient Egypt. However, Silke Roth and Toby Wilkinson point out, that the ram-hieroglyph was read differently in early times. The reading as "Ba" (meaning "soul") does not appear before the Old Kingdom period and during the two first dynasties the ram-sign was read as Khnemu (for the deity Khnum) or Ser (meaning "sheep" "ram" or "begetter"). This reading is promoted by the hieroglyph for "s" on the stela. In sum the reading on the stela had to be Seret, which means "mother sheep" or "she of the ram". It seems that the later ramesside scribes, who compiled the Annal stone (and therefore the Cairo stone inscription), had no knowledge of the older readings for the ram sign and simply read "Ba" changing Seret into Batyires.

== Tomb ==

If the lady Betrest is the same person as the lady Seret of the 1st dynasty stela, then Betrest was buried in the necropolis of king Den at Abydos. Her tomb comprises a chamber which is structurally nestled into the entrance of Den's own funerary chamber and is therefore subsidiary to his burial. Interestingly, there are two chambers fitting this description, one on the left side of the entrance and the other on the right. The two chambers differ in size but are both visibly larger than the ordinary subsidiary tombs of retainers. This peculiarity points to the privileged status that Betrest/Seret must have enjoyed during this king's reign. Indeed, only queens and royal mothers were allowed to be buried so close to the king. Unfortunately, it is not recorded in which of the two chambers the stela was found.

== References ==

Описание преобработки

Удаляем [’s], он оказался слишком частотным. В английском языке встречаются составные слова типа “easy-to-read” (здесь women-only), которые мы будем разделять на несколько токенов по дефисам, т.к. они представляют собой объединение разных слов и нет смысла учитывать их как отдельный составной токен. В английских статьях могут встречаться специальные символы других языков, как в случае Doc2 (á/...), и мы их не чистим, потому что от такого же вторжения чужеродных символов не защищены выборки ни на одном рассматриваемом языке.

4.2 Испанский

Doc1. Basílica de la Natividad de Nuestra Señora (Ratisbona).txt

Basílica de la Natividad de Nuestra Señora (Ratisbona) La Basílica de la Natividad de Nuestra Señora[1] o más formalmente Basílica de la Natividad de Nuestra Señora de la Capilla Antigua (en Alemán: Basilika Unserer Lieben Frau Zur Alten Kapelle)[2] Es el

lugar de culto católico más antiguo en Baviera y una de las iglesias más importantes de la ciudad de Ratisbona,[3] al sur de Alemania.[4] La colegiata y basílica menor es parte de una antigua abadía dedicada a la Virgen fundada por el emperador Enrique II, emperador del Sacro Imperio Romano en el 1002. Con un estilo barroco del siglo XVIII, es una de las obras maestras de la decoración rococó en Europa, en el estilo dictado por la Escuela de renombre Wessobrunn.

Una primera capilla fue construida por orden de Carlomagno en el lugar donde, según la tradición, se levantaba el templo romano dedicado a Juno. La capilla que fue construida después de la caída del Imperio Romano de Occidente, parece haber sido la más antigua de Baviera y esto le valió el apodo de la «Alte Kappelle», la Antigua Capilla.

En 1964, por sugerencia del Obispo Rudolf Graber, la Colegiata de Nuestra Señora de la Antigua Capilla fue elevada a basílica menor por el Papa Pablo VI. La Basílica de la Natividad de Nuestra Señora[1] o más formalmente Basílica de la Natividad de Nuestra Señora de la Capilla Antigua (en Alemán: Basilika Unserer Lieben Frau Zur Alten Kapelle)[2] Es el lugar de culto católico más antiguo en Baviera y una de las iglesias más importantes de la ciudad de Ratisbona,[3] al sur de Alemania.[4]

La colegiata y basílica menor es parte de una antigua abadía dedicada a la Virgen fundada por el emperador Enrique II, emperador del Sacro Imperio Romano en el 1002. Con un estilo barroco del siglo XVIII, es una de las obras maestras de la decoración rococó en Europa, en el estilo dictado por la Escuela de renombre Wessobrunn.

Una primera capilla fue construida por orden de Carlomagno en el lugar donde, según la tradición, se levantaba el templo romano dedicado a Juno. La capilla que fue construida después de la caída del Imperio Romano de Occidente, parece haber sido la más antigua de Baviera y esto le valió el apodo de la «Alte Kappelle», la Antigua Capilla.

En 1964, por sugerencia del Obispo Rudolf Graber, la Colegiata de Nuestra Señora de la Antigua Capilla fue elevada a basílica menor por el Papa Pablo VI.

== Véase también ==

Iglesia Católica

Natividad de Nuestra Señora

== Referencias ==

Doc2. Automatic Loveletter EP.txt

Automatic Loveletter EP Automatic Loveletter EP es un EP homónimo del grupo estadounidense de rock alternativo, Automatic Loveletter, lanzado el 3 de febrero de 2009 por el sello discográfico Sony BMG[2] y producido por el productor Curt Schneider. Automatic Loveletter EP es un EP homónimo del grupo estadounidense de rock alternativo, Automatic Loveletter, lanzado el 3 de febrero de 2009 por el sello discográfico Sony BMG[2] y producido por el productor Curt Schneider.

== Lista de canciones ==

== Personal ==

Juliet Simms - vocales, guitarra, guitarra rítmica

Tommy Simms - guitarra, voz secundaria

Clint Fowler - bajo

Ryan Metcalf- batería, percusión

== Referencias ==

== Enlaces externos ==

Sitio web oficial

Automatic Loveletter en PureVolume

Automatic Loveletter en YouTube

Doc3. Airest Airest (AS Airest).txt

Airest Airest (AS Airest) es una aerolínea estona que comenzó a operar en enero de 2002, operando vuelos regulares de carga y vuelos chárter de pasajeros en la ruta Tallín-Helsinki, desde 2002 para UPS como único enlace de UPS con Estonia. Airest (AS Airest) es una aerolínea estona que comenzó a operar en enero de 2002, operando vuelos regulares de carga y vuelos chárter de pasajeros en la ruta Tallín-Helsinki, desde 2002 para UPS como único enlace de UPS con Estonia.

== Airest desde 2012 ==

Airest (Airest Inc) es una compañía estona privada de carga vinculada al transporte aéreo comercial de mercancías de acuerdo con el EU-OPS Parte 1. Desde su sede central en Tallín, Airest proporciona servicios para aliados de carga del norte de Europa.

A comienzos de 2012 la Administración de Estonia de Aviación Civil (ECAA) otorgó un nuevo certificado de operaciones aérea a Airest Inc. A finales de 2012 Airest reveló un nuevo diseño de compañía, página web y boletín.

El 11 de julio de 2013 AIREST fue certificada por la ECAA como centro de mantenimiento EASA Parte 145. La organización de mantenimiento de AIREST se centra en el trabajo de mantenimiento en línea desde el SAAB 340 hasta el 800FH (incluido). Las labores de mantenimiento en línea en Tallín comenzaron a funcionar desde el 1 de agosto de 2013. En noviembre de 2014 Airest abrió su cuarta base de mantenimiento en línea en Budapest.

En septiembre de 2014 Airest firmó un acuerdo de adquisición de un quinto SAAB 340A a su flota. Tras ser registrado como ES-LSE y recibir el nombre de "Echo" efectuará vuelos en Hungría y Rumanía.

El 15 de septiembre de 2014 Airest volaba a 14 destinos con cuatro bases de mantenimiento en línea.

Los aviones de Airest están hoy volando a Arad, Baden-Baden, Billund, Budapest, Cluj-Napoca, Gotemburgo, Helsinki, Jönköping, Malmö, Stavanger, Estocolmo, Tampere, Tallín y Turku

== Flota desde 2012 ==

En noviembre de 2014, la flota de Airest se componía de: 5 Saab 340A registrados ES-LSA, ES-LSB, ES-LSC, ES-LSD, ES-LSE.

== Flota antes de 2012 ==

En agosto de 2006 la flota de Airest incluía:[1] 3 Let L-410 UVP-E registrados ES-LLA, ES-LLB y ES-LLC 1 Antonov 26B alquilado para la ruta Tallín - Helsinki, registro EW-246TG

== Referencias ==

== Enlaces externos ==

Airest Inc. Página web oficial

Boletín de Airest Inc.

Описание предобработки

Судя по испанский всему, испанский – крайне удобный язык для предобработки: ни дефисов, ни апострофов, ни гортанных смычек или других особенностей фонетики, обозначаемых чем-то напоминающим знаки препинания, в просмотренных текстах не встретилось.

4.3 Португальский

Doc1. Cerâmica plástica.txt

Cerâmica plástica Cerâmica plástica é um tipo de argila com aspecto e propriedades de gel, que tem em sua composição o polímero cloreto de polivinila (PVC). É utilizada para modelagem, criando itens de artesanato, jóias, bijuteria (bijouterias), boneco (bonecos e bonecas) e objetos de decorações. Assim, pode ser utilizada em aplicações comerciais como decoração. Atualmente, é possível encontrar arte feita com cerâmica plástica em grandes museus. Também conhecida como polymer clay em inglês.

== História ==

Baquelite, uma antiga resina, era popular com os designers e era uma forma anterior à cerâmica plástica. Contudo a base fenol do Baquelite era inflamável o que poderia gerar problemas e, por isso, **deixou-se** de ser utilizado. Dessa forma, a cerâmica plástica foi formulada para possivelmente tomar o lugar do baquelite. A formulação atraiu a atenção da alemã Käthe Kruse, criadora de bonecas, em 1939. Enquanto não era viável **utilizá-la** em sua fábrica, Kruse deixou sua filha Sophie, cujo apelido era "Fifi brincar e modelar a massa. A fórmula, mais tarde, foi vendida para Eberhardt Faber que disponibilizou no mercado com o nome "FIMO"(Fifi's MOdeling compound - Composto de modelagem da Fifi, em português).

== Referências ==

== Links Externos ==

International Polymer Clay Association

Doc2. Rondo Cameron.txt

Rondo Cameron Rondo Cameron (1925 – 1 de janeiro de 2001) foi um professor norte-americano de história econômica. Era nativo do Texas. Formou-se em Yale (1948) e recebeu o grau de pH.d. na Universidade de Chicago (1952). Lecionou na Universidade do Wisconsin-Madison a partir de 1952. Em 1969 foi para a Universidade de Emory, onde foi professor da Universidade de Kenan até se aposentar como professor emérito em 1993. Foi presidente da International Economic History Association (Sylla, 2001). É bem conhecido dos alunos de primeiro ano do curso de história da Universidade de Leicester pelo seu lendário livro de 1989 - A Concise Economic History of the World: From Paleolithic Times to the Present ("Uma história concisa da economia do mundo: do paleolítico até ao presente"). Segundo o prefácio, o livro teve muitos a ser amadurecido, com passagens num capítulo de uma palestra introdutória do curso de graduação em Yale em 1951. Rondo Cameron (1925 – 1 de janeiro de 2001) foi um professor **norte-americano** de história econômica. Era nativo do Texas. Formou-se em Yale (1948) e recebeu o grau de pH.d. na Universidade de Chicago (1952). Lecionou na Universidade do Wisconsin-Madison a partir de 1952. Em 1969 foi para a Universidade de Emory, onde foi professor da Universidade de Kenan até se aposentar como professor emérito em 1993. Foi presidente da International Economic History Association (Sylla, 2001). É bem conhecido dos alunos de primeiro ano do curso de história da Universidade de Leicester pelo seu lendário livro de 1989 - A Concise Economic History of the World: From Paleolithic Times to the Present ("Uma história concisa da economia do mundo: do paleolítico até ao presente"). Segundo o prefácio, o livro teve muitos a ser amadurecido, com passagens num capítulo de uma palestra introdutória do curso de graduação em Yale em 1951.

== Publicações selecionadas ==

Jerome Blum, Rondo Cameron, and Thomas G. Barnes. (1970). The European World. Little, Brown. Rondo Cameron et al. (1975). Civilizations: Western and World.

- Rondo Cameron (1966). France and Economic Development of Europe.
 (1967). Banking in the Early Stages of Industrialization.
 (1989). A Concise Economic History of the World: From Paleolithic Times to the Present, Oxford.
 Rondo Cameron and Larry Neal (2003, 4th ed.) A Concise Economic History of the World: From Paleolithic Times to the Present, 480 pp., including annotated bibliography, Oxford.
 == Ligações externas ==
 Rondo Cameron 1925-2001. Um tributo, Emory University Department of History Newsletter, de agosto de 2001.
 Richard Sylla (2001) EH.N: Rondo Cameron, 1925-2001. Dados biográficos e avaliação profissional.

Doc3. Carlos Alberto Parreira.txt

Carlos Alberto Parreira Carlos Alberto Gomes Parreira (Rio de Janeiro, 27 de fevereiro de 1943) é um **ex-treinador** de futebol brasileiro que participou de Copas do Mundo por cinco seleções diferentes (Arábia Saudita, Brasil, Emirados Árabes, Kuwait e África do Sul), sendo também o técnico brasileiro que mais dirigiu seleções de outros países, seis ao todo, em dez passagens.

Parreira é conhecido nacional e mundialmente pelas conquistas da Copa do Mundo de 1994, ao lado de Zagallo, do Campeonato Brasileiro de 1984 com o Fluminense e a Copa do Brasil de 2002, com o Corinthians, entre outros títulos nacionais conquistados no exterior. Carlos Alberto Gomes Parreira (Rio de Janeiro, 27 de fevereiro de 1943) é um ex-treinador de futebol brasileiro que participou de Copas do Mundo por cinco seleções diferentes (Arábia Saudita, Brasil, Emirados Árabes, Kuwait e África do Sul), sendo também o técnico brasileiro que mais dirigiu seleções de outros países, seis ao todo, em dez passagens. Parreira é conhecido nacional e mundialmente pelas conquistas da Copa do Mundo de 1994, ao lado de Zagallo, do Campeonato Brasileiro de 1984 com o Fluminense e a Copa do Brasil de 2002, com o Corinthians, entre outros títulos nacionais conquistados no exterior.

== Carreira ==

Formado em Educação Física pela Escola Nacional de Educação Física e Desportos, no Rio de Janeiro, em 1966, começou sua carreira como Preparador Físico do São Cristóvão de Futebol e Regatas. Foi um dos preparadores físicos da Seleção Brasileira na Copa do Mundo de 1970, quando a seleção conquistou o tricampeonato mundial. Como técnico, Parreira trabalhou em uma série de clubes pelo Brasil e pelo mundo. Suas passagens mais marcantes foram no Fluminense (seu time do coração), em 1984, quando o clube conquistou seu segundo título do Campeonato Brasileiro e em 1999, quando ajudou na recuperação do clube com a conquista da Série C do Campeonato Brasileiro. Além do clube carioca, Parreira também fez história no Corinthians em 2002, quando foi campeão da Copa do Brasil e da última edição do Torneio Rio-São Paulo, e no Bragantino, levando o modesto clube do interior paulista à final do Campeonato Brasileiro de 1991, perdida para o São Paulo de Telê Santana.

Ao todo, integrou a comissão técnica de sete clubes brasileiros, boa parte do tempo no Rio de Janeiro. Mesmo com passagens vitoriosas por essas equipes, foi após a conquista da Copa de 1994, com a Seleção Brasileira, que Parreira obteve reconhecimento e prestígio internacionais, que renderam-lhe diversos convites e propostas de trabalho no exterior, abrindo portas num mercado até então muito pouco explorado. E foi aí que sua carreira decolou. Dirigiu Valencia CF (Espanha), Fenerbahçe (Turquia) e MetroStars na recém-iniciada Major League Soccer.

Mas o maior sucesso foi com seleções nacionais. Conquistou vários torneios asiáticos com

as seleções do Kuwait e da Arábia Saudita, além da Copa América 2004 e a Copa do Mundo de 1994 pela Seleção Brasileira. Atingiu ainda a notável façanha de classificar cinco seleções nacionais diferentes e disputar seis Copas do Mundo: Kuwait - 1982; Emirados Árabes - 1990; Brasil - 1994 e 2006; Arábia Saudita - 1998; África do Sul - 2010, igualando assim o recorde anterior (já que Bora Milutinović levou cinco países à competição), são resultados bastante expressivos para um profissional até hoje bastante contestado pelos críticos do esporte e, principalmente, pela exigente torcida brasileira.

Muitos faziam reservas ao estilo mais cauteloso e burocrático de Parreira, além de tê-lo considerado excessivamente teórico e esquemático. Há os que defendem que as equipes de Parreira frequentemente eram caracterizadas pelo bom toque de bola e pelo envolvimento do adversário em seu jogo. Inspirado no lema de Sepp Herberger (antigo técnico alemão) que diz que futebol é atacar e defender com a máxima eficiência, Parreira ficou conhecido pela frase "O Gol É apenas um detalhe mostrando a intensa visão do jogo defensivo.

Até 17 abril de 2008, Parreira comandou a seleção da África do Sul, anfitriã da próxima Copa do Mundo. Entretanto, alegando problemas particulares, o treinador deixa a seleção **sul-africana** sem obter resultados satisfatórios. Em seu lugar, assume o também brasileiro Joel Santana, recém saído do Flamengo. Em 6 de março de 2009, Parreira acerta a sua volta ao Fluminense (seu clube de coração), após a saída de René Simões. No entanto, em 13 de julho do mesmo ano, o técnico tetracampeão mundial pela seleção brasileira foi demitido por conta dos resultados inconsistentes no Campeonato Brasileiro.

Описание предобработки

Ничего, что мешало бы определению языка, не обнаружено, и мы можем пренебречь несколькими частицами, которые пишутся через дефис и появляются только в этих позициях.

4.4 Indonезийский

Doc1. Halme basifusca.txt

Halme basifusca Halme basifusca adalah spesies kumbang tanduk panjang yang tergolong familia Cerambycidae. Spesies ini juga merupakan bagian dari genus Halme, ordo Coleoptera, kelas Insecta, filum Arthropoda, dan kingdom Animalia.

Larva kumbang ini biasanya mengebor ke dalam kayu dan dapat menyebabkan kerusakan pada batang kayu hidup atau kayu yang telah ditebang. Halme basifusca adalah spesies kumbang tanduk panjang yang tergolong familia Cerambycidae. Spesies ini juga merupakan bagian dari genus Halme, ordo Coleoptera, kelas Insecta, filum Arthropoda, dan kingdom Animalia.

Larva kumbang ini biasanya mengebor ke dalam kayu dan dapat menyebabkan kerusakan pada batang kayu hidup atau kayu yang telah ditebang.

== Referensi ==

TITAN: Cerambycidae database. Tavakilian G., 25 Mei 2009.

Nugini.txt

Monarki Papua Nugini Templat:Politik Papua Nugini Monarki Papua Nugini adalah sebuah sistem pemerintahan dimana seorang penguasa monarki warisan menjadi penguasa berdaulat dan kepala negara Papua Nugini. Penguasa monarki saat ini, sejak 16 September 1975, adalah Ratu Elizabeth II. Meskipun penguasa berdaulatnya berbagi kesetaraan dengan lima belas negara merdeka lainnya pada Kerajaan Persemakmuran, monarki di setiap negara tersebut terpisah dan memiliki hukum yang berbeda. Akibatnya, penguasa monarki saat ininya secara resmi digelar Ratu Papua Nugini dan, dalam kapasitas ini, ia, permaisurinya,

dan anggota keluarga Kerajaan lainnya memegang fungsi pribadi dan publik secara domestik dan luar negeri sebagai perwakilan negara Papua Nugini. Namun, Ratu merupakan satu-satunya anggota Keluarga Kerajaan dengan peran konstitusional. Ratu biasanya tinggal di Britania Raya dan, meskipun beberapa kekuasaan ia pegang sendiri, sebagian besar tugas seremonial dan pemerintahan kerajaan di Papua Nugini diserahkan kepada perwakilan Ratu, gubernur-jenderal.

Tanggung jawab penguasa berdaulat, dan gubernur-jenderal, di bawah konstitusi Papua Nugini, meliputi membentuk dan membubarkan parlemen, mengadakan pemilihan, dan melantik pemerintahan. Templat:Politik Papua Nugini Monarki Papua Nugini adalah sebuah sistem pemerintahan dimana seorang penguasa monarki warisan menjadi penguasa berdaulat dan kepala negara Papua Nugini. Penguasa monarki saat ini, sejak 16 September 1975, adalah Ratu Elizabeth II. Meskipun penguasa berdaulatnya berbagi kesetaraan dengan lima belas negara merdeka lainnya pada Kerajaan Persemakmuran, monarki di setiap negara tersebut terpisah dan memiliki hukum yang berbeda. Akibatnya, penguasa monarki saat

ininya secara resmi digelar Ratu Papua Nugini dan, dalam kapasitas ini, ia, permaisurinya, dan anggota keluarga Kerajaan lainnya memegang fungsi pribadi dan publik secara domestik dan luar negeri sebagai perwakilan negara Papua Nugini. Namun, Ratu merupakan satu-satunya anggota Keluarga Kerajaan dengan peran konstitusional. Ratu biasanya tinggal di Britania Raya dan, meskipun beberapa kekuasaan ia pegang sendiri, sebagian besar tugas seremonial dan pemerintahan kerajaan di Papua Nugini diserahkan kepada perwakilan Ratu, gubernur-jenderal. Tanggung jawab penguasa berdaulat, dan gubernur-jenderal, di bawah konstitusi Papua Nugini, meliputi membentuk dan membubarkan parlemen, mengadakan pemilihan, dan melantik pemerintahan.

== Referensi ==

Templat:Topik Papua Nugini Templat:Kerajaan Persemakmuran

Templat:Elizabeth II

Описание предобработки

Необходимо удалить все слова, написанные не латинскими буквами, учитывать наличие гортанной смычки, которая отмечается апострофом на конце слов: ta', ga' итд. В целом индонезийский язык прост для токенизации.

4.5 Литовский

Doc1. Fotodiodas.txt

Fotodiodas Fotodiodas – diodas, savo sandara pritaikytas atlikti šviesos daviklio funkcijas. Šviesos kvantai suteikia elektronams papildomos energijos, padėdami įveikti uždarytą pn sandūrą. Todėl stebimi du efektai: Apšvietus padidėja „atvirkščia“ (tamsoje beveik srovės nepraleidžiančia) kryptimi įjungtu diodu tekanti srovė. Diodas pradeda dirbti kaip saulės baterija, verčianti šviesos energiją elektros energija.

Šie du efektai pasireiškia ir paprastame diode, jei jo korpusas pagamintas iš šviesai laidžios medžiagos. Specialiais gamybos metodais abu efektai gali būti labai sustiprinti. Fotodiodas tinka akiai nematomiems infraraudoniesiems, ultravioletiniams bei rentgeno spinduliams aptikti, jei tik jo korpusas šiuos spindulius praleidžia. Jei reikia ypatingo jautrumo (paskiriems fotonams skaičiuoti, fotodiodas darbo metu atšaldomas iki labai žemų temperatūrų. Skirtingai nuo fotorezistoriaus, fotodiodas į šviesą reaguoja labai greitai. Pavyzdžiui, elektrinėje gitaroje jų greičio pakanka registruoti stygos virpesiams. Tinka jie ir įvairiai informacijai, valdymo signalams priimti (signalų šaltinis šiuo atveju paprastai būna šviesos diodas). Fotodiodas yra labiausiai paplitęs nuotolinio valdymo pultelio komandas priimančias daviklis. Fotodiodas – diodas, savo sandara pritaikytas atlikti šviesos daviklio funkcijas. Šviesos kvantai suteikia elektronams papildomos energijos, padėdami įveikti uždarytą pn sandūrą. Todėl stebimi du

efektai: Apšvietus padidėja „atvirkščia“ (tamsoje beveik srovės nepraleidžiančia) kryptimi įjungtu diodu tekanti srovė. Diodas pradeda dirbti kaip saulės baterija, verčianti šviesos energiją elektros energija.

Šie du efektai pasireiškia ir paprastame diode, jei jo korpusas pagamintas iš šviesai laidžios medžiagos. Specialiais gamybos metodais abu efektai gali būti labai sustiprinti. Fotodiodas tinka akiai nematomiems infraraudoniesiems, ultravioletiniams bei rentgeno spinduliams aptikti, jei tik jo korpusas šiuos spindulius praleidžia. Jei reikia ypatingo jautrumo (paskiriems fotonams skaičiuoti, fotodiodas darbo metu atšaldomas iki labai žemų temperatūrų. Skirtingai nuo fotorezistoriaus, fotodiodas į šviesą reaguoja labai greitai. Pavyzdžiui, elektroninėje gitaroje jų greičio pakanka registruoti stygos virpesiams. Tinka jie ir įvairiai informacijai, valdymo signalams priimti (signalų šaltinis šiuo atveju paprastai būna šviesos diodas). Fotodiodas yra labiausiai paplitęs nuotolinio valdymo pultelio komandas priimančias daviklis.

Doc2. Mantas Armalis.txt.

Mantas Armalis Mantas Armalis (g. 1992 m. rugsėjo 6 d. Plungėje) – ledo ritulio vartininkas, fotomodelis. Mantas Armalis (g. 1992 m. rugsėjo 6 d. Plungėje) – ledo ritulio vartininkas, fotomodelis.

== Biografija ==

Gimė 1992 m. rugsėjo 6 d. Plungėje. Būdamas 4 mėnesių kartu su tėvais, kurie buvo orientavimosi sporto meistrai, išvyko gyventi į Švediją. Ledo ritulininko karjerą pradėjo 2007–2008 m. Djurgården U16 komandoje. 2008–2011 m. žaidė už Mora IK jaunimo U18 ir U20 komandas. 2011–2012 m. sezone pradėjo žaisti pagrindinėje Mora IK komandoje, tačiau vėliau buvo paskolintas Söderhamn/Ljusne komandai, kuri žaidė C1 divizione. Šiame divizione ledo ritulininkas debiutavo 2011 m. spalio 5 d. prieš Lindlövens IF komandą, kuriai pralaimėjo 5 – 3. Vėliau buvo paskolintas Borlänge HF komandai, kurioje sužaidė vienas varžybas. 2012–2013 m. sezoną pradėjo Tranås AIF komandoje. 2008 m. debiutavo Lietuvos U18 ledo ritulio rinktinėje, nuo 2010 m. U20 rinktinėje, o 2012 m. buvo pripažintas geriausiu Ukrainoje vykusio U20 II diviziono turnyro A grupės vartininku.

== Nuorodos ==

18-metis vartininkas Mantas Armalis įsitvirtino ne tik Lietuvos ledo ritulio rinktinėje, bet ir Švedijoje;

Ledo ritulininkas M.Armalis – pirmasis lietuvis, žengęs Milano mados savaitės podiumu; Eurohockey Profilis (ang.);

hockeyDB Profilis (ang.);

Elite Prospects Profilis (ang.);

Susiję straipsniai www.lrytas.lt.

== Šaltiniai ==

Doc3. Grigorio de Chirico.txt

Giorgio de Chirico Džordžas de Kirikas (it. Giorgio de Chirico, 1888 m. liepos 10 d., Volas, Graikija – 1978 m. lapkričio 19 d., Roma) – XX a. italų tapytojas, iliustratorius, dizaineris, memuaristas. Italų metafizinės tapybos vienas iš pradininkų. Džordžas de Kirikas (it. Giorgio de Chirico, 1888 m. liepos 10 d., Volas, Graikija – 1978 m. lapkričio 19 d., Roma) – XX a. italų tapytojas, iliustratorius, dizaineris, memuaristas. Italų metafizinės tapybos vienas iš pradininkų.

== Biografija ==

Džordžas de Kirikas gimė 1888 m. liepos 10 d. Volo mieste Graikijoje. Jo tėvai, geležinkelietis inžinierius Evaristo de Chirico ir motina Gemma Cervetto, buvo italai. 1900 m. Dž. de Kirikas pradėjo mokytis Atėnų politechnikos institute ir vakarais lankė nuogo kūno piešimo

užsiėmimus. 1906 m. jis persikėlė į Miuncheną, kur lankė Vaizduojamojo meno akademiją, susidomėjo Arnold Böcklin ir Max Klinger daile, A. Šopenhauerio ir F. Nyčės filosofija. 1909 m. Dž. de Kirikas persikėlė į Milaną, 1910 m. į Florenciją, 1911 m. į Paryžių. Florencijoje Dž. de Kirikas pradėjo kurti serijas peizažų, pavyzdžiui, „Rudens popietės enigma“, kur vaizduojami ilgi, nenatūralūs objektų šešėliai tuščiose miestų erdvėse su nematomų objektų ryškia šviesa. Jo kūriniai buvo eksponuoti 1912 ir 1913 m. Paryžiaus Rudens Salone ir 1913 ir 1914 m. Nepriklausomųjų parodoje. Dž. de Kirikas buvo nuolatinis G. Apolinero pobūvių lankytojas, kur susitiko Konstantiną Brankušį, Andrė Dereną, Maksą Žakobą ir kt.

Pirmojo pasaulinio karo metu Dž. Kirikas buvo pakviestas tarnauti kariuomenėje, grįžo į Italiją, kur Feraroje 1916 m. susitiko Filippo de Pisis ir 1917 m. Carlo Carrà, su kuriais suformavo taip vadinamą „metafizikų mokyklą“. 1918 m. Dž. de Kirikas apsigyveno Romoje. 1918-19 m. žiemą įvyko pirmoji jo personalinė paroda Romos Casa d'Arte Bragaglia. Jis priklausė Valori Plastici menininkų ir estetų grupei. 1920–1924 m. Dž. de Kirikas gyveno tai Romoje, tai Florencijoje. 1921 m. įvyko jo personalinė paroda Dailės galerijoje Milane ir pirmą kartą jo kūriniai 1924 m. pristatyti Venecijos bienalėje. Nuo 1925 m. dailininkas gyveno Paryžiuje, eksponavo L'Effort Moderne (1925), Polio Gijomo (1926) ir Jeanne Bucher (1927) galerijose. 1928 m. įvyko jo personalinės parodos Londone Arthur Tooth galerijoje ir Niujorke Valentino galerijoje. 1929 m. Dž. de Kirikas kūrė kostiumus ir scenos dekoracijas S. Diagilevo trupės pasirodymams ir buvo publikuotas jo poetinis romanas Hebdomeros. Sekančiais metais dailininkas kūrė teatro ir operos pastatymams, jo kūriniai buvo eksponuoti JAV, Kanadoje, Japonijoje. 1945 m. publikuotas pirmas tomas jo „Mano gyvenimo atsiminimų“. Nuo 1944 m. Dž. de Kirikas gyveno Romoje. Ankstyvoji avangardinė Dž. Kiriko kūryba įtakojė siurrealistų dailę. Kartu su Carlo Carrà ir Giorgio Morandi apie 1917 m. Dž. de Kirikas išvystė tapybos stilių, vadinamą „metafizine tapyba“. Po 1930 m. Dž. de Kirikas labiau susidomėjo klasikinės dailės formomis, atsiribojo nuo savo ankstyvųjų avangardinių kūrinių ir pradėjo tapyti labiau akademine maniera portretus, natūrmortus, aktus. Jis sukūrė 66 litografijas G. Apolinero „Kaligramų“ iliustravimui. Po karo dalyvavo „antimoderniųjų“ tapytojų parodose, iliustravo A. Mandzonio „Sužadėtinių“ ir „Iliados“ itališkus leidimus. 1960-ųjų pabaigoje pradėjo kurti bronzines skulptūras. Dž. de Kirikas mirė 1978 m. lapkričio 20 d. Romoje. Jo palaikai 1992 m. perlaidoti San Francesco a Ripa bažnyčioje.

1986 m. Romoje įsteigtas Džordžo ir Izos de Kirikų fondas, skirtas saugoti dailininko kūrybinį palikimą. 1998 m. atsidarė Dž. de Kiriko namas-muziejus Palazzetto dei Borgognoni (31 Piazza di Spagna) Romoje, kur dailininkas gyveno paskutinius 30 metų su antrąja žmona Isabella Pakszwer Far (m. 1990 m.). Jis 1930 m. buvo vedęs rusų baleriną ir aktorę Raisą Gurevič Krol, su kuria išsiskyrė 1931 m. jau tuo metu pradėjęs susitikinėti su Isabella Pakszwer Far, kurią vedė 1946 m.

== Kūrinių galerijos ==

Vikiteka

wikiart.org, kūriniai chronologine tvarka

Dž. de Kiriko kūriniai Niujorko šiuolaikinio meno muziejaus svetainėje

Dž. de Kiriko kūriniai Barnso fonde, Filadelfija

Dž. de Kiriko fondo paveikslai

Dž. de Kiriko fondo skulptūros

Dž. de Kiriko fondo piešiniai

== Šaltiniai ==

guggenheim.org/

Encyclopaedia Britannica

Išplėstinė biografija Džordžo ir Izos de Kirikų fondo svetainėje

Описание предобработки

Можно токенизировать по пробелам, но есть сложности с обозначениями года и

месяца в датах: g. 1992 m. rugsėjo 6 d., есть заимствования/слова из других языков: Casa d'Arte Bragaglia. И если заимствованные названия встречаются редко, то слово год "metai" и слово "diena" день, которые сокращаются до m. d., если их не убрать, станут самыми частотными словами в литовском словаре.

4.6 Польский

Doc1.Ivica Dačić.txt

Ivica Dačić Ivica Dačić, сыр. Ивица Дачић (ur. 1 stycznia 1966 w Prizrenie) – serbski polityk, premier Serbii od 27 lipca 2012 do 27 kwietnia 2014. Od 2006 przewodniczący Socjalistycznej Partii Serbii, od 2008 do 2014 minister spraw wewnętrznych, od 2014 pierwszy wicepremier i minister spraw zagranicznych, od maja do czerwca 2017 p.o. premiera Serbii.

Ivica Dačić, (ur. 1 stycznia 1966 w Prizrenie) – serbski polityk, premier Serbii od 27 lipca 2012 do 27 kwietnia 2014. Od 2006 przewodniczący Socjalistycznej Partii Serbii, od 2008 do 2014 minister spraw wewnętrznych, od 2014 pierwszy wicepremier i minister spraw zagranicznych, od maja do czerwca 2017 p.o. premiera Serbii.

== Žyciorys ==

Ivica Dačić urodził się w 1966 w Prizrenie na terenie Kosowa. Wychowywał się w miejscowości Žitorađa. Jego ojciec, Desimir, był policjantem, podczas gdy matka, Jelisaveta, zajmowała się domem. Uczęszczał do szkoły w Niszu. W 1989 ukończył studia dziennikarskie na Wydziale Nauk Politycznych Uniwersytetu w Belgradzie. Ivica Dačić jest żonaty, ma dwoje dzieci. Deklaruje znajomość języka angielskiego i rosyjskiego. W działalność polityczną zaangażował się na początku lat 90. W 1990 został przewodniczącym belgradzkiej młodzieżówki socjalistycznej. Rok później oficjalnie wstąpił w szeregi Socjalistycznej Partii Serbii (SPS) kierowanej przez Slobodana Miloševicia. Na początku lat 90. był redaktorem ukazującego się przez pewien czas czasopisma tej partii „Epoha”. W czasie wyborów w 1992 oraz w 1993 kierował sprawami informacji i propagandy. W 1992 wszedł w skład kierownictwa SPS. W tym samym roku został również wybrany na deputowanego do Izby Obywateli Zgromadzenia Federalnego Jugosławii. W latach 1992–2002 pełnił funkcję rzecznika Socjalistycznej Partii Serbii, zyskując przydomek „Małego Słobo”. W 2000 stanął na czele struktur partii w Belgradzie, a także został wybrany na wiceprzewodniczącego SPS.

Po odsunięciu w październiku 2000 od władzy prezydenta Slobodana Miloševicia objął urząd ministra informacji w przejściowym rządzie Milomira Minicia, w którym pozostał do stycznia 2001. W 2003 objął mandat deputowanego do Zgromadzenia Narodowego Republiki Serbii i stanął na czele klubu parlamentarnego SPS. W wyborach prezydenckich z 13 czerwca 2004 zajął piąte miejsce, uzyskawszy 4,04% głosów poparcia.

3 grudnia 2006 został wybrany na stanowisko przewodniczącego Socjalistycznej Partii Serbii. W 2007 został prezesem klubu piłkarskiego KK Partizan Belgrad, w przeszłości zajmował również stanowisko wiceprzewodniczącego Jugosłowiańskiego Komitetu Olimpijskiego. W wyborach parlamentarnych z 11 maja 2008 koalicja SPS pod jego przywództwem zajęła czwarte miejsce, zdobywając 20 mandatów w parlamencie. Po wyborach zawarł koalicję rządową ze zwycięską Partią Demokratyczną prezydenta Borisa Tadića. W konsekwencji 8 lipca 2008 objął urząd wicepremiera i ministra spraw wewnętrznych w gabinecie Mirka Cvetkovicia. Był również serbskim członkiem Zgromadzenia Parlamentarnego Rady Europy.

W wyborach parlamentarnych z 6 maja 2012 koalicja skupiona wokół Socjalistycznej Partii Serbii zajęła trzecie miejsce, zdobywając 44 spośród 250 mandatów. W przeprowadzonej tego samego dnia pierwszej turze wyborów prezydenckich Ivica Dačić również zajął trzecie miejsce, uzyskawszy 14,9% głosów.

W wyniku rozmów koalicyjnych lider socjalistów zdecydował się na opuszczenie koalicji z Partią Demokratyczną, której lider prezydent Boris Tadić przegrał 20 maja 2012 drugą

ture wyborów prezydenckich z Tomislavem Nikolicem, przewodniczącym Serbskiej Partii Postępowej. W czerwcu SPS zawiązała koalicję ze zwycięską Serbską Partią Postępową. 28 czerwca 2012 Ivica Dačić został desygnowany przez nowego prezydenta na stanowisko premiera rządu. 27 lipca 2012 zaprzysiężony został nowy gabinet, w którym zachował również urząd ministra spraw wewnętrznych.

W przedterminowych wyborach w 2014 Ivica Dačić ponownie kierował listą SPS-PUPS-JS, uzyskując mandat poselski. 27 kwietnia 2014 na urzędzie premiera zastąpił go Aleksandar Vučić z Serbskiej Partii Postępowej. Ivica Dačić pozostał w nowym gabinecie jako pierwszy wicepremier oraz minister spraw zagranicznych.

W wyniku kolejnych wyborów w 2016 lider socjalistów również wszedł w skład Skupštiny, będąc liderem zorganizowanej przez swoją partię koalicji. Aleksandar Vučić początkowo nie zamierzał kontynuować koalicji z SPS, jednakże sojusz rządowy został utrzymany. W sierpniu 2016 ogłoszono ponowną nominację Ivicy Dačića na pierwszego wicepremiera i ministra spraw zagranicznych w drugim gabinecie dotychczasowego premiera.

31 maja 2017, po objęciu przez Aleksandara Vučića stanowiska prezydenta, został pełniącym obowiązki premiera. Pełnił tę funkcję do 29 czerwca 2017, gdy zaprzysiężony został rząd Any Brnabić. W gabinecie tym Ivica Dačić pozostał na dotychczasowych stanowiskach.

== Przypisy ==

Doc2. O duchu praw.txt

O duchu praw O duchu praw (fr. *De l'esprit des lois*) – traktat filozoficzny Monteskiusza, wydany po raz pierwszy w Genewie w 1748. Jest to jedno z fundamentalnych dzieł francuskiego oświecenia, a zawarty w nim postulat trójpodziału władzy stanowi podstawę porządku ustrojowego większości współczesnych państw demokratycznych.

Dzieło to, uznawane za najważniejsze w dorobku filozofa, stanowi zwieńczenie jego pracy naukowej i sumę życiowych doświadczeń.

Ze względu na postępującą utratę wzroku treść traktatu była dyktowana sekretarzom. O duchu praw (fr. *De l'esprit des lois*) – traktat filozoficzny Monteskiusza, wydany po raz pierwszy w Genewie w 1748. Jest to jedno z fundamentalnych dzieł francuskiego oświecenia, a zawarty w nim postulat trójpodziału władzy stanowi podstawę porządku ustrojowego większości współczesnych państw demokratycznych.

Dzieło to, uznawane za najważniejsze w dorobku filozofa, stanowi zwieńczenie jego pracy naukowej i sumę życiowych doświadczeń.

Ze względu na postępującą utratę wzroku treść traktatu była dyktowana sekretarzom.

== Tło historyczne ==

Dzieło Monteskiusza powstało w czasie, gdy na francuskim tronie zasiadali przedstawiciele dynastii Burbonów.

Sukcesy długiego panowania Ludwika XIV (1643–1715) – wewnętrzny spokój i polityka merkantylizmu zapewniały monarchii olbrzymie dochody, a tłumienie opozycji wobec króla (szczególnie uchylenie w 1685 edyktu nantejskiego o tolerancji religijnej doprowadziło do masowej emigracji francuskich protestantów) oraz zręczne rozgrywanie antagonizmów między szlachtą i mieszczaństwem – umocniły monarchię absolutną we Francji, pozostawiając w rękach króla nieograniczoną władzę nad największą europejską potęgą owego czasu.

Jednak zaangażowanie Ludwika XV (1715–1774) w wojny terytorialne z Wielką Brytanią i z Habsburgami o sukcesję austriacką (1741–1748), rosnące koszty utrzymania licznych zamorskich kolonii (m.in. Kanada, Luizjana), a także przepych dworu i pogłębiająca się nędza chłopów doprowadziły do załamania budżetu i kryzysu królestwa. Zarówno despotyzm ówczesnej władzy, jak i niepokoje społeczne stanowią tło rozważań O duchu praw. Na język polski traktat Monteskiusza przełożył Tadeusz Boy-Żeleński.

== Przypisy ==

== Bibliografia ==

Norman Davis: Europa. Kraków: Wydawnictwo ZNAK, 2003.

Judith N. Shklar: Montesquieu. Oxford – New York: Oxford University Press, 1987.

Doc3. Nauka w Wietnamie.txt

Nauka w Wietnamie Główną instytucją naukową w Wietnamie jest Unia Towarzystw Naukowych i Technologicznych Wietnamu w Hanoi (założona w 1983), która pełni funkcję akademii nauk i jest organem doradczym rządu. Organem rządowym ds. nauki jest Komitet Rady Ministrów do spraw Nauki, któremu podlega Narodowe Centrum Badań Naukowych Wietnamu, obejmujące 12 instytutów badawczych z zakresu nauk ścisłych oraz instytut badań atomowych.

Poszczególnym ministerstwom i departamentom podlegają 72 instytuty, prowadzące badania podstawowe, stosowane i wdrożeniowe. W Wietnamie działają 4 uniwersytety: w Hanoi (założony w 1956, kształcący 31 tys. studentów - 2003), w Cn Th (założony w 1966, kształcący 14 tys. studentów - 2003) i w Ho Chi Minh (założony w 1977, kształcący 8 tys. studentów - 2003). Ponadto są 2 politechniki: w Hanoi (założona w 1966, kształcąca 14 tys. studentów - 2003) i w Ho Chi Minh (założona w 1962, kształcąca 16,6 tys. studentów - 2003) oraz 40 wyższych szkół zawodowych - technicznych, medycznych, ekonomicznych i innych. Główną instytucją naukową w Wietnamie jest Unia Towarzystw Naukowych i Technologicznych Wietnamu w Hanoi (założona w 1983), która pełni funkcję akademii nauk i jest organem doradczym rządu. Organem rządowym ds. nauki jest Komitet Rady Ministrów do spraw Nauki, któremu podlega Narodowe Centrum Badań Naukowych Wietnamu, obejmujące 12 instytutów badawczych z zakresu nauk ścisłych oraz instytut badań atomowych.

Poszczególnym ministerstwom i departamentom podlegają 72 instytuty, prowadzące badania podstawowe, stosowane i wdrożeniowe. W Wietnamie działają 4 uniwersytety: w Hanoi (założony w 1956, kształcący 31 tys. studentów - 2003), w Cn Th (założony w 1966, kształcący 14 tys. studentów - 2003) i w Ho Chi Minh (założony w 1977, kształcący 8 tys. studentów - 2003). Ponadto są 2 politechniki: w Hanoi (założona w 1966, kształcąca 14 tys. studentów - 2003) i w Ho Chi Minh (założona w 1962, kształcąca 16,6 tys. studentów - 2003) oraz 40 wyższych szkół zawodowych - technicznych, medycznych, ekonomicznych i innych.

Описание предобработки

Польский язык хорошо токенизируется по пробелам. Сложных случаев для токенизации не было найдено, только слово "Ссылка" тоже нужно задать специальное для этого языка (przypisy).

4.7 Чешский

Doc1. Affero General Public License.txt

Affero General Public License Affero General Public License, často zkracováno jako Affero GPL nebo AGPL (a někdy neformálně též Affero licence) je označení pro dvě rozdílné, avšak historicky související licence svobodného softwaru: (1) Affero General Public License, verze 1 (publikována společností Affero, Inc. v březnu 2002 a založená na GNU General Public License, verze 2 (GPLv2)), a (2) GNU Affero General Public License, verze 3 (publikována Free Software Foundation v listopadu 2007 a odvozená z GNU General Public License, verze 3 (GPLv3)).

Obě verze AGPL byly navrženy pro uzavření skuliny v běžné GPL, týkající se poskytovatelů aplikačních služeb (ASP). Každá z verzí se liší od GNU GPL, na které je založena, tím, že přidává další podmínku ohledně užití softwaru přes síť. Tato podmínka vyžaduje, aby

každý ze síťových uživatelů díla licencovaného pod GPL (typicky webové aplikace) měl přístup ke kompletním zdrojovým kódům. Free Software Foundation doporučuje, aby se uvažovalo o použití GNU AGPLv3 u každého softwaru, který se bude obvykle používat přes síť. Affero General Public License, často zkracováno jako Affero GPL nebo AGPL (a někdy neformálně též Affero licence) je označení pro dvě rozdílné, avšak historicky související licence svobodného softwaru: (1) Affero General Public License, verze 1 (publikována společností Affero, Inc. v březnu 2002 a založená na GNU General Public License, verze 2 (GPLv2)), a (2) GNU Affero General Public License, verze 3 (publikována Free Software Foundation v listopadu 2007 a odvozená z GNU General Public License, verze 3 (GPLv3)).

Obě verze AGPL byly navrženy pro uzavření skuliny v běžné GPL, týkající se poskytovatelů aplikačních služeb (ASP). Každá z verzí se liší od GNU GPL, na které je založena, tím, že přidává další podmínku ohledně užití softwaru přes síť. Tato podmínka vyžaduje, aby každý ze síťových uživatelů díla licencovaného pod GPL (typicky webové aplikace) měl přístup ke kompletním zdrojovým kódům. Free Software Foundation doporučuje, aby se uvažovalo o použití GNU AGPLv3 u každého softwaru, který se bude obvykle používat přes síť.

== Historie ==

V roce 2000, během vývoje obchodního modelu pro e-learning a e-slужby, setkal se Henry Poole v Amsterdamu se Richardem Stallmanem a diskutovali tam o ASP skulině v GPLv2. Během následujících měsíců řešili, jak se vypořádat s tímto problémem. V roce 2001 Poole založil Affero, Inc. (firmu zaměřenou na webové služby) a potřeboval licenci, která by vyžadovala distribuci kódu těmi organizacemi, které použijí kód společnosti Affero k vytvoření odvozených webových služeb. Tehdy Poole kontaktoval Bradleyho M. Kuhna a Ebena Moglena z Free Software Foundation, aby mu poradili s novou licencí, která by uzavřela ASP skulinu v GPLv2.

Někdy koncem února 2002 Kuhn navrhl, na základě ideje programu, který vypisuje svůj vlastní zdrojový kód, aby byla GPLv2 doplněna sekci 2(d), která by vyžadovala, aby odvozená díla měla funkci pro stažení zdrojového kódu, určenou k poskytnutí úplného odpovídajícího zdrojového kódu. Kuhn argumentoval tím, že v sekci 2(c) je precedent pro takový požadavek, kde se vyžaduje uchování určitých funkcí všemi, kdo budou dílo šířit nebo upravovat.

Moglen a Kuhn sepsali text nově navržené sekce 2(d) a poskytli ho Poolovi, který si potom vyžádal a obdržel souhlas FSF k publikování odvozeniny GPLv2 pro tyto účely. V březnu 2002 publikovala firma Affero, Inc. původní Affero General Public License (AGPLv1) pro použití s projektem Affero a dala tuto licenci k dispozici k použití jinými vývojáři vytvářejícími software jako službu.

FSF uvažovala zahrnout zvláštní podmínku pro AGPLv1 do GPLv3, ale jednomyslně se rozhodla vydat samostatnou licenci, téměř identickou s GPLv3, ale obsahující podmínku s podobným účelem a efektem, jako má sekce 2(d) v AGPLv1. Nová licence byla označena jako GNU Affero General Public License, přičemž zahrnutí názvu Affero indikuje těsnou historickou vazbu na AGPLv1. GNU AGPL dostala číslo verze 3, vzhledem ke své blízkosti ke GPL téže verze, aktuální verze GNU Affero General Public License se tedy často zkracuje jako AGPLv3. Finální verze AGPLv3 byla vydána FSF 19. listopadu 2007. Prvním známým softwarovým systémem vydaným pod AGPLv3 (21. listopadu 2007) je Stet. Mezi široce používané projekty pod licencí AGPLv3 patří petiční web předsedy vlády Velké Británie.

== Kompatibilita s GPL ==

Obě verze AGPL, podobně jako odpovídající verze GNU GPL, na kterých jsou založeny, jsou silně copyleftové licence. Podle úsudku FSF činí dodatečný požadavek v sekci 2(d) licence AGPLv1 tuto nekompatibilní s jinak téměř identickou GPLv2. Tedy nelze distribuovat dílo vytvořené kombinací komponent, které mají každá jednu z těchto licencí.

V kontrastu k tomu, GPLv3 a AGPLv3 obsahují klauzule (v sekci 13 každé z licencí),

které společně zajišťují vzájemnou kompatibilitu těchto dvou licencí. Tyto klauzule explicitně povolují poskytování díla vytvořeného slinkováním kódu licencovaného pod jednou z licencí s kódem licencovaným pod licencí druhou. V tomto případě se copyleft každé licence oslabí, aby byla povolena distribuce takových kombinací. K vytvoření cesty pro upgrade z AGPLv1 na AGPLv3 od FSF publikovala firma Affero licenci Affero General Public License verze 2, která je toliko přechodovou licencí, umožňující příjemcům softwaru licencovaného pod "AGPLv1 nebo libovolnou novější verzí publikovanou Affero, Inc." distribuovat software nebo odvozená díla pod AGPLv3.

== Související články ==

GNU General Public License

GNU Lesser General Public License

GNU Free Documentation License

== Reference ==

V tomto článku byl použit překlad textu z článku Affero General Public License na anglické Wikipedii.

== Externí odkazy ==

GNU Affero General Public License Version 3

GPLv3, software jako služba a verze 2 licence Affero GPL

Affero General Public License Version 1

AGPL Version 1 FAQ

Free Software Foundation podporuje AGPL - tisková zpráva

Free Software Foundation GNU AGPL v3 - tisková zpráva

Doc 2. Baskicko-islandský pidžin

Baskicko-islandský pidžin Baskicko-islandský pidžin je pidžin, kterým se mluvilo v sedmnáctém století na Islandu a na lodích v atlantském oceánu mezi baskickými rybáři, mořeplavci a lovci velryb. Jedná se o baskičtinu s mnoha slovy z islandštiny, nizozemštiny, angličtiny, francouzštiny, španělštiny a němčiny. Na Islandu se používal na poloostrově Vestfirir. Baskičtí mořeplavci byli velmi významní, existují i teorie že objevili Ameriku a na Island se proto dostávali velmi často. Kryštof Kolumbus dokonce před objevením Ameriky navštívil Island a zřejmě zde objevil mimo jiné i baskické záznamy. Baskicko-islandský pidžin je pidžin, kterým se mluvilo v sedmnáctém století na Islandu a na lodích v atlantském oceánu mezi baskickými rybáři, mořeplavci a lovci velryb. Jedná se o baskičtinu s mnoha slovy z islandštiny, nizozemštiny, angličtiny, francouzštiny, španělštiny a němčiny. Na Islandu se používal na poloostrově Vestfirir. Baskičtí mořeplavci byli velmi významní, existují i teorie že objevili Ameriku a na Island se proto dostávali velmi často. Kryštof Kolumbus dokonce před objevením Ameriky navštívil Island a zřejmě zde objevil mimo jiné i baskické záznamy.

== Ukázky baskicko-islandského pidžinu ==

V následující tabulce jsou ukázky baskicko-islandského pidžinu a pro srovnání je uvedena baskičtina a islandština. Ukázky jsou z velrybářské knihy Vocabula Biscaica, uložené v Rejkjavíku.

== Příklady některých slov z baskicko-islandského pidžinu ==

==== Slova baskického původu ====

atorra, z atorra 'košile'

berria, z berria 'nový'

berrua, z beroa 'horký'

bustana, z buztana 'ocas'

eta, z eta 'a'

galsardia, z galtzerdia 'ponožka'

gissuna, z gizona 'muž'
locaria, z lokarria 'krajka'
sagarduna, z sagardoa 'cider'
ser, z zer 'co'
sumbatt, z zenbat 'kolik'
travala, z travaillatu, 'pracovat', sice přejato z baskičtiny, ale románského původu
usnia, z esnea 'mléko'

==== Slova germánského původu ====

===== Slova z angličtiny =====

for mi, z for me 'pro mě', lze použít ve více významech

for ju, z for you 'pro tebe', lze použít ve více významech

===== Slova z ostatních germánských jazyků =====

for, není známo, ze kterého z germánských jazyků toto slovo pochází, ale v germánských jazycích for znamená pro.

cavinit, z gar nichts, což znamená v nizozemštině a němčině vůbec nic.

==== Slova románského původu ====

===== Slova ze španělštiny =====

cammisola, z camisola 'košile'

fenicha, z fornicar 'smilnit'

trucka, z trocar 'vyměnit'

===== Slova z francouzštiny nebo španělštiny =====

mala, z mal 'špatný'

eliza, z église 'kostel'

==== Ano ====

Ano se baskicko-islandským pidžinem řeklo bai (z baskického bai) nebo vi (z francouzského vý, v moderní francouzštině psáno oui)

== Reference ==

V tomto článku byly použity překlady textů z článků Basque–Icelandic pidgin na anglické Wikipedii a Pidgin vasco-islandés na španělské Wikipedii.

Doc 3. Alex Miller.txt

Alex Miller (politik) Alex Miller (rusky: Алекс Миллер, Aleks Miller) je izraelský politik a poslanec Knesetu za stranu Jisra'el Bejtejnu. Alex Miller (rusky: Алекс Миллер, Aleks Miller) je izraelský politik a poslanec Knesetu za stranu Jisra'el Bejtejnu.

== Biografie ==

Narodil se 4. dubna 1977 v Moskvě, v tehdejším Sovětském svazu, dnes Rusko. V roce 1992 přesídlil do Izraele. Získal bakalářský titul z pedagogiky a magisterský z veřejné politiky. Žije ve městě Ariel na Západním břehu Jordánu. Je ženatý, má jedno dítě. Hovoří hebrejsky, anglicky a rusky.

== Politická dráha ==

Působil jako předseda studentského svazu na škole Ort Academic College v Tel Avivu, byl místopředsedou celostátního studentského svazu a předsedá mládežnické organizaci strany Jisra'el Bejtejnu.

Do Knesetu nastoupil po volbách roku 2006, ve kterých kandidoval za stranu Jisra'el Bejtejnu. Ve funkčním období 2006-2009 zastával post člena parlamentního výboru pro televizi a rozhlas, výboru pro ekonomické záležitosti, výboru pro ústavu, právo a spravedlnost, výboru pro vzdělávání, kulturu a sport, výboru pro vědu a technologie, výboru pro práva dětí a člena parlamentní vyšetřovací komise pro odposlechy. Mandát obhájil ve volbách roku 2009.

Od roku 2009 působí jako místopředseda Knesetu. Dále je členem parlamentního finančního výboru a předsedou výboru pro vzdělávání, kulturu a sport.

== Odkazy ==

==== Reference ====

==== Externí odkazy ====

(anglicky) Kneset – Alex Miller

Описание предобработки

В чешском языке по большей части нет отличительных особенностей, требующих отдельных правил при токенизации.

Есть возвратная частица *se* (или *si*), которую хотелось бы соединить с глаголом в один токен, т.к. без возвратной частицы глагол имеет другое значение, но *se* может встречаться не только в качестве возвратной частицы, поэтому ввести регулярное выражение для объединения *se* и глагола не получится.

Проблемы при токенизации могут быть связаны с наличием статей с большим количеством слов на другом языке (например, статья про баскско-исландский пиджин на чешском языке). Однако в общей массе текстов эти слова не должны оказаться наиболее частотными, а избавиться от них в процессе предобработки не представляется возможным.