

Jaká data se používají ve strojovém překladu

Ondřej Bojar

bojar@ufal.mff.cuni.cz

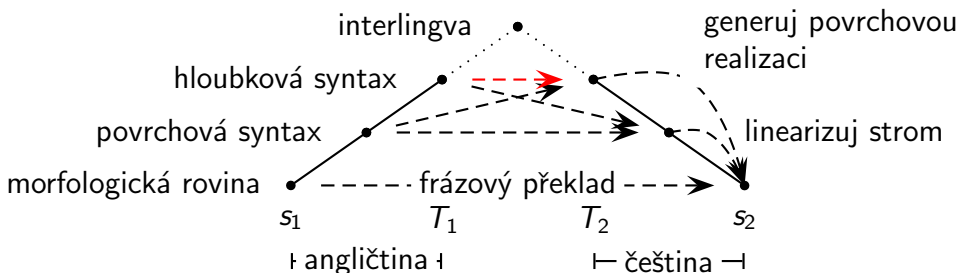
Ústav formální a aplikované lingvistiky
MFF UK

ELRC Workshop, 14. duben 2016

Osnova

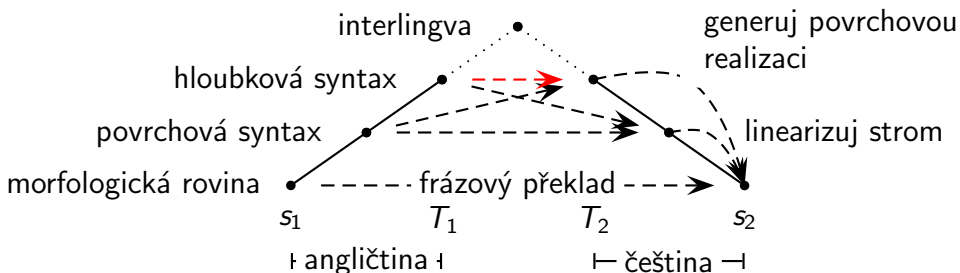
- ▶ Typy dat ve strojovém překladu.
- ▶ Kolik je potřeba?
- ▶ Absolutní minimum: paralelní korpusy.
 - ▶ Získávání paralelních textů z webu.
- ▶ Sdílení zdrojů.

Přístupy ke strojovému překladu



- ▶ Pravidlový vs. statistický přístup:
 - ▶ Pravidlové systémy píší lingvisté-programátoři.
 - ▶ Statistické systémy se naučí samy podle dat.

Přístupy ke strojovému překladu



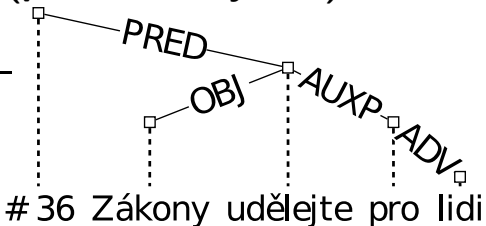
- ▶ Pravidlový vs. statistický přístup:
 - ▶ Pravidlové systémy píší lingvisté-programátoři.
 - ▶ Statistické systémy se **naučí samy podle dat**.

Formální popis češtiny \rightsquigarrow Ruční anotace

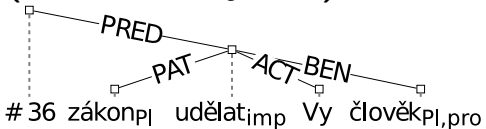
Morfologická rovina:

Slovo	Lema	Morfologická značka
zákony	zákon	NNIP1----A----
zákony	zákon	NNIP4----A----
zákony	zákon	NNIP5----A----
zákony	zákon	NNIP7----A----
udělejte	udělat	Vi-P---2--A----
udělejte	udělat	Vi-P---3--A---4
pro	pro-1	RR--4-----
lidi	člověk	NNMP1----A----
lidi	člověk	NNMP4----A----
lidi	člověk	NNMP5----A----

Analytická rovina (povrchová syntax):



Tektogramatická rovina (hloubková syntax):



Lingvistická data

- ▶ **Korpusy** jsou (velké) sbírky textů:
 - ▶ Texty typicky označované nebo včetně větných rozborů.
Pražský závislostní korpus (PDT): 1.5 mil. slov.
Pražský čj-aj závislostní korpus (PCEDT): 50 tis. vět.
 - ▶ Některé vícejazyčné: CzEng (15 mil. vět, 220 mil. slov, odpovídá ~50 metrům knih, ty tvoří však jen čtvrtinu).
- ▶ **Slovníky** strojově čitelné:
 - ▶ Morfologický slovník říká, že *kočka* je české slovo a *kočke* ne.
 - ▶ Valenční slovník říká, že:
Rodiče přijali Petra. → je správně
Rodiče přijeli Petra. → není správně
 - ▶ Slovník subjektivity obsahuje hodnotící výrazy.

⇒ Lze využít v programech (pravidlových i statistických).

Data po frázový překlad (Moses, MT@EC)

- ▶ Překladové slovníky méně vhodné.
 - ▶ Překlad potřebuje vidět slova v kontextu.
- ▶ Terminologie a glosáře méně vhodné.
 - ▶ Současné metody mohou vynutit překlad jen velice tupě.
- ▶ Zásadní je paralelní korpus.
- ▶ Pro kvalitu důležité, aby texty byly z dané domény.
- ▶ Systém memoruje až desetice slov.
 - ⇒ Dlouhé termíny zachovány přesně.
 - ⇒ České tvarosloví a volný slovosled situaci komplikuje.

Paralelní korpus

GENESIS

The Story of Creation

1 In the beginning, when God created the universe,² the earth was formless and desolate. The raging ocean that covered everything was engulfed in total darkness, and the Spirit of God was moving over the water.³ Then God commanded, "Let there be light" – and light appeared.⁴ God was pleased with what he saw. Then he separated the light from the darkness,⁵ and he named the light "Day" and the darkness "Night". Evening passed and morning came – that was the first day.

⁶⁻⁷ Then God commanded, "Let there be a dome to divide the water and to keep it in two separate places" – and it was done. So God made a dome, and it separated the water under it from the water above it.⁸ He named the dome "Sky". Evening passed and morning came – that was the second day.

⁹ Then God commanded, "Let the water below the sky come together in one place, so that the land will appear" – and it was done.¹⁰ He named the land "Earth", and the water which had come together he named "Sea". And God was pleased with what he saw.¹¹ Then he commanded, "Let the earth produce all kinds of plants, those that bear grain and those that bear fruit" – and it was done.¹² So the earth produced all kinds of plants, and God was pleased with what he saw.¹³ Evening passed and morning

GENÈSE

Dieu crée l'univers et l'humanité

1 Au commencement Dieu créa le ciel et la terre.

² La terre était sans forme et vide, et l'obscurité couvrait l'océan primitif. Le souffle de Dieu se déplaçait à la surface de l'eau.³ Alors Dieu dit: "Que la lumière paraisse!" et la lumière parut.⁴ Dieu constata que la lumière était une bonne chose, et il sépara la lumière de l'obscurité.⁵ Dieu nomma la lumière jour et l'obscurité nuit. Le soir vint, puis le matin; ce fut la première journée.

⁶ Dieu dit encore: "Qu'il y ait une voûte, pour séparer les eaux en deux masses!"⁷ Et cela se réalisa. Dieu fit ainsi la voûte qui sépare les eaux d'en bas de celles d'en haut.⁸ Il nomma cette voûte ciel. Le soir vint, puis le matin; ce fut la seconde journée.

⁹ Dieu dit encore: "Que les eaux qui sont au-dessous du ciel se rassemblent en un lieu unique pour que le continent paraisse!" Et cela se réalisa.¹⁰ Dieu nomma le continent terre et la masse des eaux mer, et il constata que c'était une bonne chose.¹¹ Dieu dit alors: "Que la terre produise de la végétation: des herbes produisant leur semence, et des arbres fruitiers dont chaque espèce porte ses propres graines!" Et cela se réalisa.¹² La terre fit pousser de la végétation: des herbes produisant leur semence

⇒ Automatické zarovnání po větách

In my dream , there was a sycamore growing out of the ruins of the sacristy , and I was told that , if I dug at the roots of the sycamore , I would find a hidden treasure . But I ' m not so stupid as to cross an entire desert just because of a recurrent dream . " And they disappeared . The boy stood up shakily , and looked once more at the Pyramids . " It is I who dared to do so , " said the boy . This man looked exactly the same , except that now the roles were reversed . " It is I who dared to do so , " he

अपने सपने में मुझे एक गुलर का पेड़ दिखाई देता था और मुझे लगता था कि अगर मैं उस गुलर की जड़ें खोद डालूं तो मुझे छिपा हुआ खजाना मिल जाएगा । मगर मैं तुम्हारी तरह इतना बेवकूफ नहीं हूँ कि महज बार - बार आने वाले एक सपने के कारण पूरे रेगिस्तान को पार करूं । वे लोग , उसके बाद वहां से चले गए । लड़का लड़खड़ाता हुआ किसी तरह खड़ा हो गया । <s>एक बार फिर उसने पिरामिडों को देखा । " यह जुर्रत मैंने की थी , " लड़के ने कहा । <s>उसे सेंटियागो मातामोरोस की वह प्रतिमा याद आई जिसमें वह घोड़े पर सवार था और उसके घोड़े के खुर्शों में कितने ही नास्तिक कुचले हुए पड़े थे । यह घुड़सवार भी बिलकुल वैसा ही था । यह बात और थी कि इनके किरदार बदले हुए थे । " मैंने ही ऐसा करने का साहस किया था , " लड़के ने दोहराया और अपनी गर्दन तलवार का वार सहने के लिए झुका दी । ' जिंदगी ने भी हमेशा मेरे साथ अच्छा बर्ताव किया । '

Automatické zářivky povětách

In my dream, there was a sycamore growing out of the ruins of the sanctuary, and I was told that, if I dug at the roots of the sycamore, I would find a hidden treasure.
But I'm not so stupid as to cross an entire desert just because of a recurrent dream.

The boy stood up shakily, and looked once more at the Pyramids.
And they disappeared.

"It is I who dared to do so," said the boy.
This man looked exactly the same, except that now his miles were reversed.
"It is I who dared to do so," he repeated, and he lowered his head to receive a blow from the sword.
"Life was good to me," the man said.

"When you appeared in my dream, I felt that all my efforts had been rewarded, because my son's poems will be read by men for generations to come."

I don't want anything for myself.
But any father would be proud of the fame achieved by one whom he had cared for as a child, and educated as he grew up.
"We're two very different things."
"That's not true," the boy said.

"I learned the alchemist's secrets in my travels, the stars, and everything created in the universe."
I have inside me the winds, the deserts, the oceans, the stars, and everything created in the universe.
You'll learn to love the desert, and you'll get to know every one of the fifty thousand palms.
You'll watch them as they grow, demonstrating how the world is always changing.

And you'll get better and better at understanding omens, because the desert is the best teacher there is.
"Sometime during the second year, you'll remember about the treasure."
The omens will appear incoherently at first, and you'll try to ignore them.

But you know that I'm not going to go to Mecca, just as you know that you're not going to buy your sheep.
"Who told you that?" asked the boy, startled.
"Maktub" said the old crystal merchant.

And he gave the boy his blessing.
The boy went to his room and packed his belongings.

As he was leaving, he saw, in the corner of the room, his old shepherd's pouch.
I want to see the greatness of Allah, the chief said, with respect.

But he made a mental note of the names of the two men who had expressed their fear.
"Tell them about the armies that are approaching."
"They are men of the desert, and the men of the desert are used to dealing with omens."

As soon as the wind stopped, he was going to remove them from their commands, because true men of the desert are not afraid.
"Well, then, they probably already know."
"They're not concerned with that right now."

They believe that if they have to know about something Allah wants them to know, someone will tell them about it.
"It has happened many times before."

"What about one tenth of my treasure?"
The old man looked disappointed.
"If you start out by promising what you don't even have yet, you'll lose your desire to work toward getting it."
The boy told him that he had already promised to give one-tenth of his treasure to the Gypsy.
"Gypsies are experts at getting people to do that," suggested the old man.
The older continued, "When the pharaoh dreamed of cows that were thin and cows that were fat, this man I'm speaking of rescued Egypt from famine."

His name was Joseph, and he was probably young by your age.
He passed, and his eyes were still unfriendly.
"We always observe the Tradition."

"Everyone on earth has a treasure that awaits him," his heart said.
"We speak of them only to children. Later, we simply let life proceed, in its own direction, toward its own fate."
But he didn't need to worry about that right now.
"We always observe the Tradition."

What he had to be concerned about was his treasure, and how he was going to go about getting it.
The sale of his sheep had left him with enough money in his pouch, and the boy knew that in money there was magic.
Before long, maybe in just a few days, he would be at the Pyramids.
An old man, with a breastplate of gold, would lead him to acquire his sheep.

And when two such people encounter each other, the past and the future become unimportant.
It is the hand that evokes love, and creates a twin soul for every person in the world.
Without such love, one's dreams would have no meaning.
"Maktub," thought the boy.

"Hmm..." said the old man, looking at all sides of the book, as if it were some strange object.
"The Englishman shook the boy, and the boy's eyes were wide open."
"Hmm..." said the old man, looking at all sides of the book, as if it were some strange object.

„Snoke stane v tisku, jak jsem se v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal.“

„Snoke stane v tisku, jak jsem se v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal.“

„Snoke stane v tisku, jak jsem se v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal.“

„Snoke stane v tisku, jak jsem se v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal.“

„Snoke stane v tisku, jak jsem se v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal.“

„Snoke stane v tisku, jak jsem se v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal.“

„Snoke stane v tisku, jak jsem se v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal.“

„Snoke stane v tisku, jak jsem se v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal.“

„Snoke stane v tisku, jak jsem se v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal.“

„Snoke stane v tisku, jak jsem se v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal.“

„Snoke stane v tisku, jak jsem se v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal.“

„Snoke stane v tisku, jak jsem se v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal.“

„Snoke stane v tisku, jak jsem se v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal.“

„Snoke stane v tisku, jak jsem se v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal.“

„Snoke stane v tisku, jak jsem se v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal.“

„Snoke stane v tisku, jak jsem se v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal.“

„Snoke stane v tisku, jak jsem se v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal.“

„Snoke stane v tisku, jak jsem se v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal.“

„Snoke stane v tisku, jak jsem se v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal, a on mi ukazuje, ze jsem v tom vyznamal.“

Zarovnaný paralelní korpus (1658)

☞:☼:(2):☼:☞

Invitatio.

Einleitung.



M. Veni, Puer!
disce Sapere.

P. Quid hoc est,
Supere?

M. Omnia,
quæ necessaria,
rectè intelligere,
rectè agere,
rectè eloqui.

P. Quis me
hoc docebit?

M. Ego,
cum DEO.

L. Komm her/ Knab!
lerne Weißheit.

S. Was ist das/
Weißheit?

L. Alles/
was nöthig ist/
recht verstehen/
recht thun/
recht ausreden.

S. Wer wird mich
das lehren?

L. Ich/
mit GOTT.

⇒ Automatické zarovnání po slovech

what	█	-	-	-	-	-	-	-	-	-	-	-	-
is	-	█	-	-	-	-	-	-	-	-	-	-	-
the	-	-	█	-	-	-	-	-	-	-	-	-	-
length	-	-	█	-	-	-	-	-	-	-	-	-	-
of	-	-	█	-	-	-	-	-	-	-	-	-	-
the	-	-	█	-	-	-	-	-	-	-	-	-	-
vector	-	-	-	█	-	-	-	-	-	-	-	-	-
that	-	-	-	-	█	-	-	-	-	-	-	-	-
is	-	-	-	-	-	█	-	-	-	-	-	-	-
the	-	-	-	-	-	-	█	-	-	-	-	-	-
sum	-	-	-	-	-	-	█	-	-	-	-	-	-
of	-	-	-	-	-	-	█	-	-	-	-	-	-
all	-	-	-	-	-	-	-	█	-	-	-	-	-
of	-	-	-	-	-	-	-	█	-	-	-	-	-
these	-	-	-	-	-	-	-	-	█	-	-	-	-
?	-	-	-	-	-	-	-	-	-	█	-	-	-

jaká vektoru součet
je , že všech
délka je těchto
?

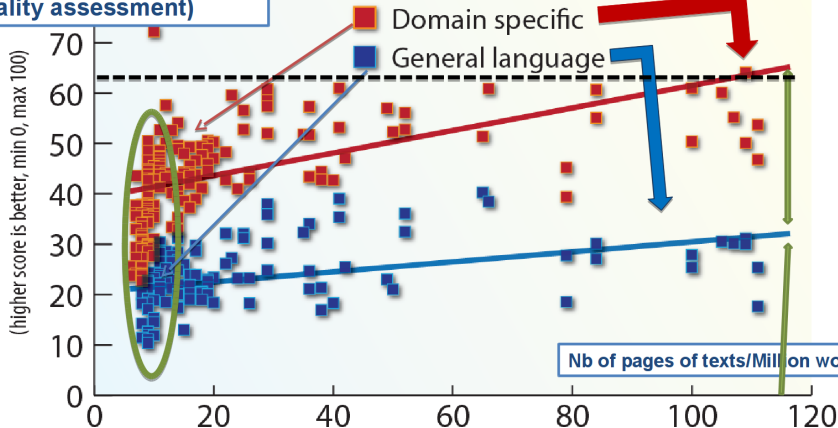
Kolik textů je potřeba?

Kolik textů je potřeba?

Víc.

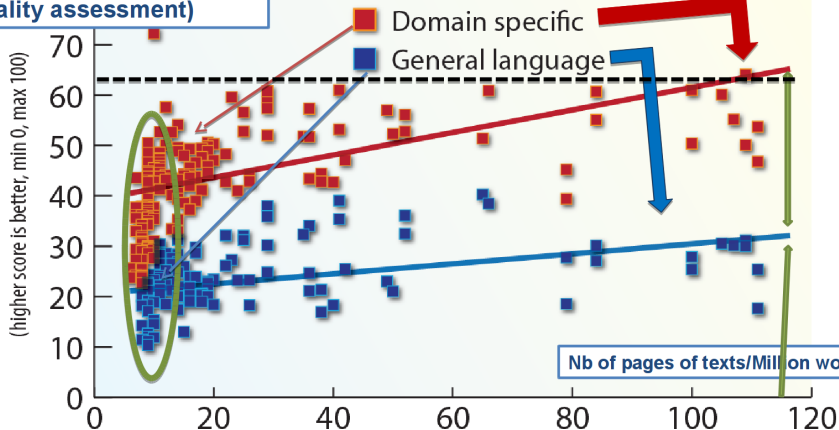
Víc textů \Rightarrow vyšší kvalita

A Commonly used measure
(quality assessment)



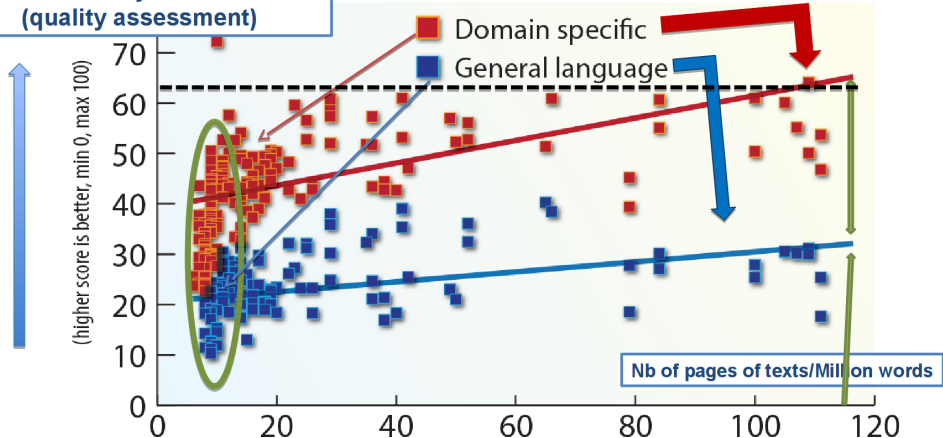
Texty v doméně \Rightarrow rovnou lepší výstup

A Commonly used measure
(quality assessment)



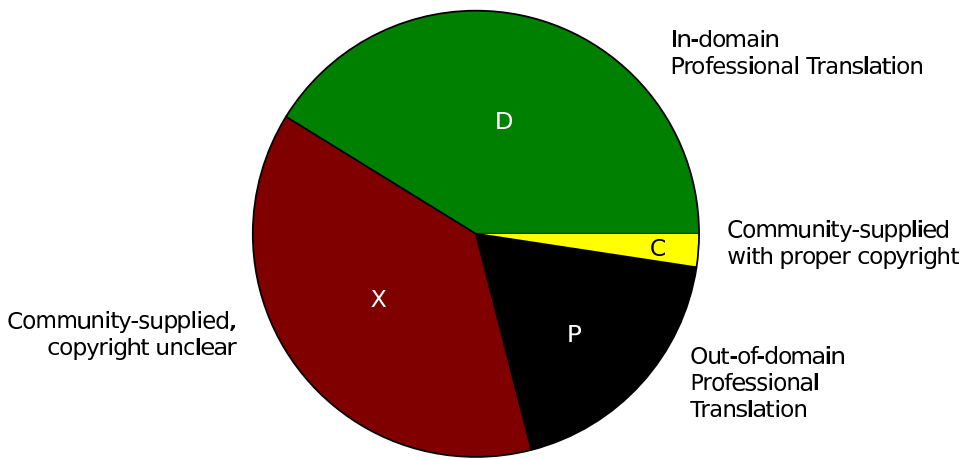
Texty v doméně \Rightarrow rychlejší zlepšení

A Commonly used measure
(quality assessment)

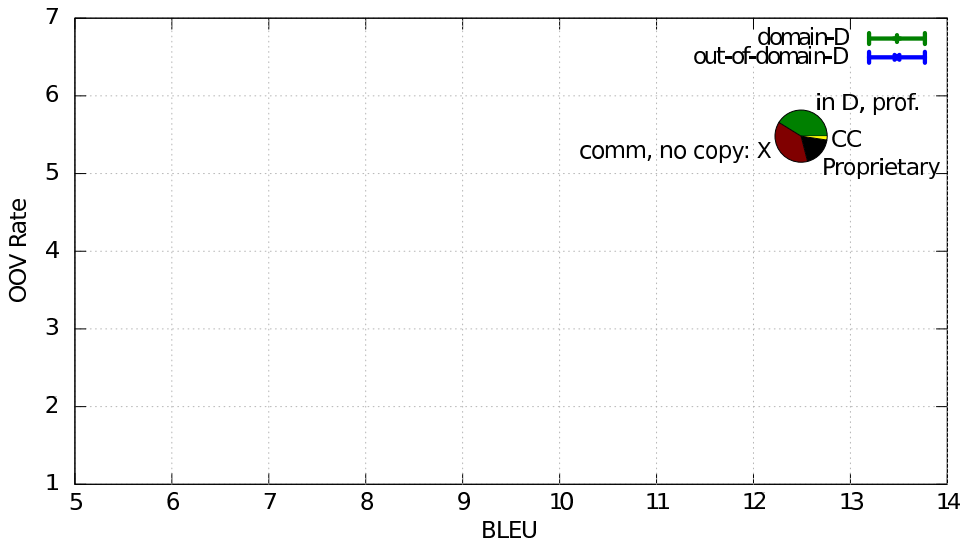


Data pro aj→čj v roce 2008

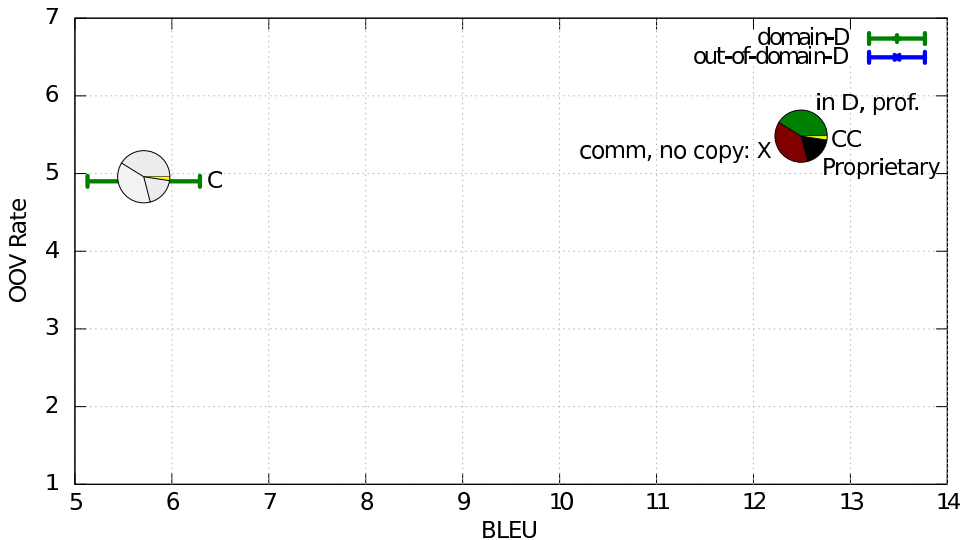
Training Data Composition



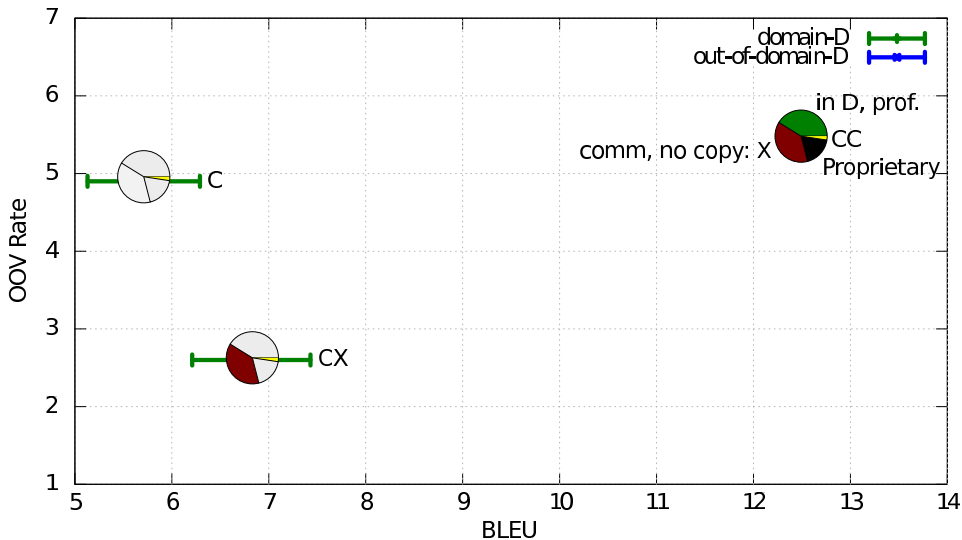
Vliv dat na kvalitu a neznámá slova



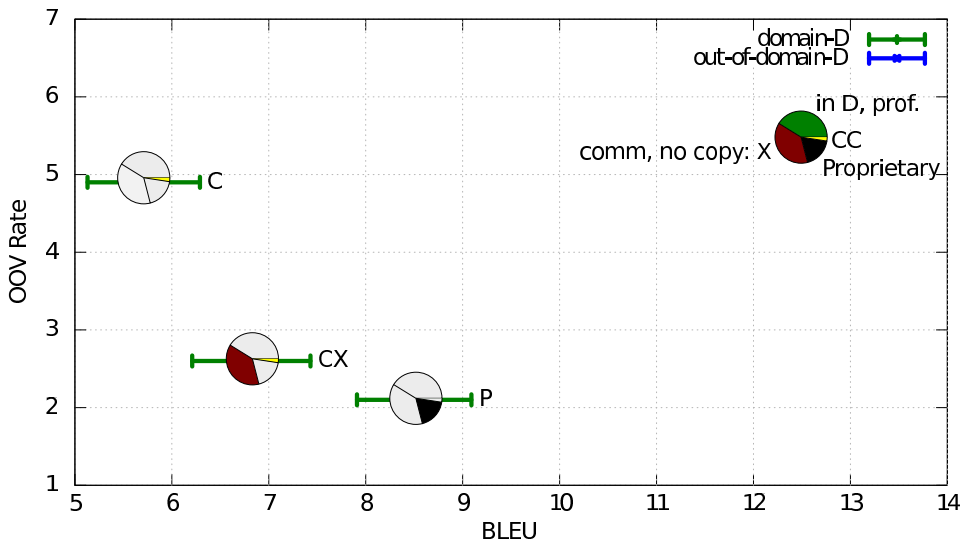
Komunitní data mimo doménu



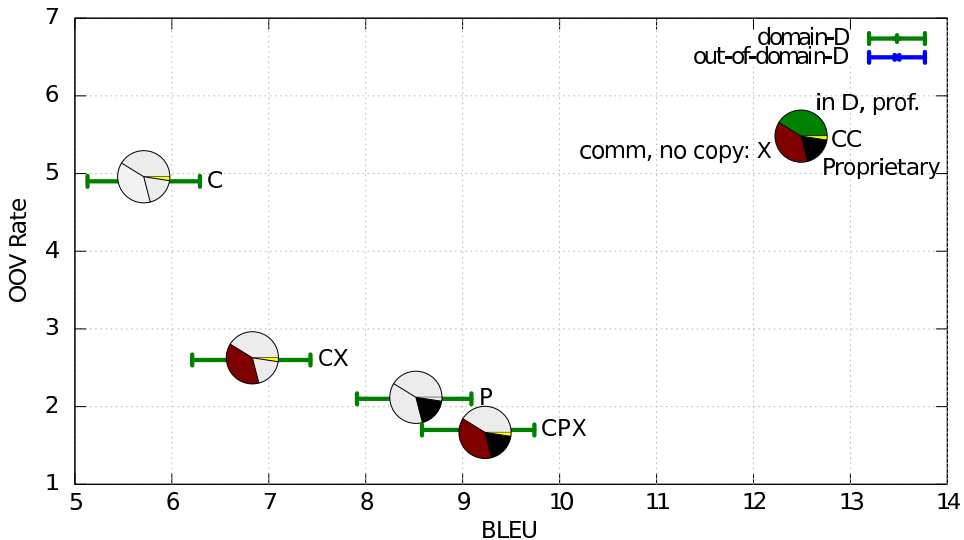
Komunitní data mimo doménu



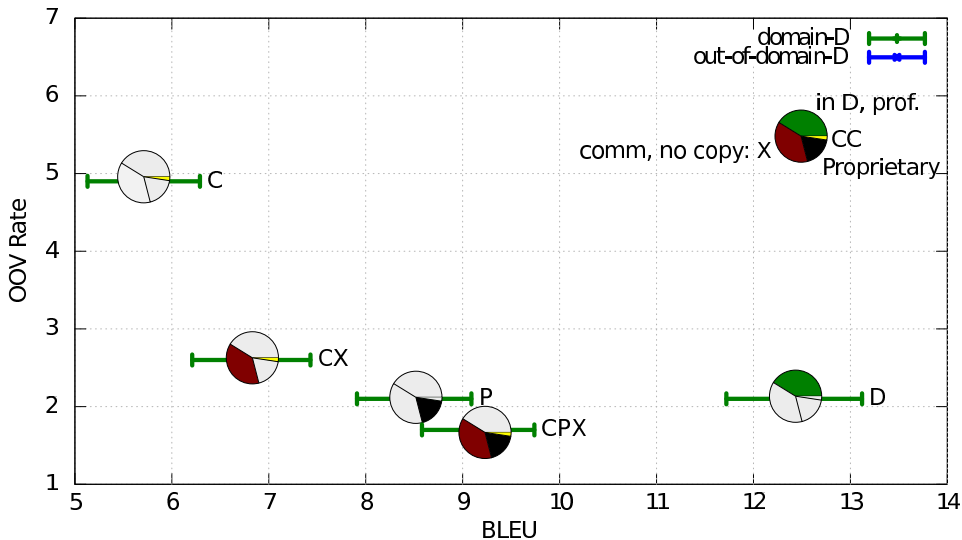
Profesionální překlady mimo doménu



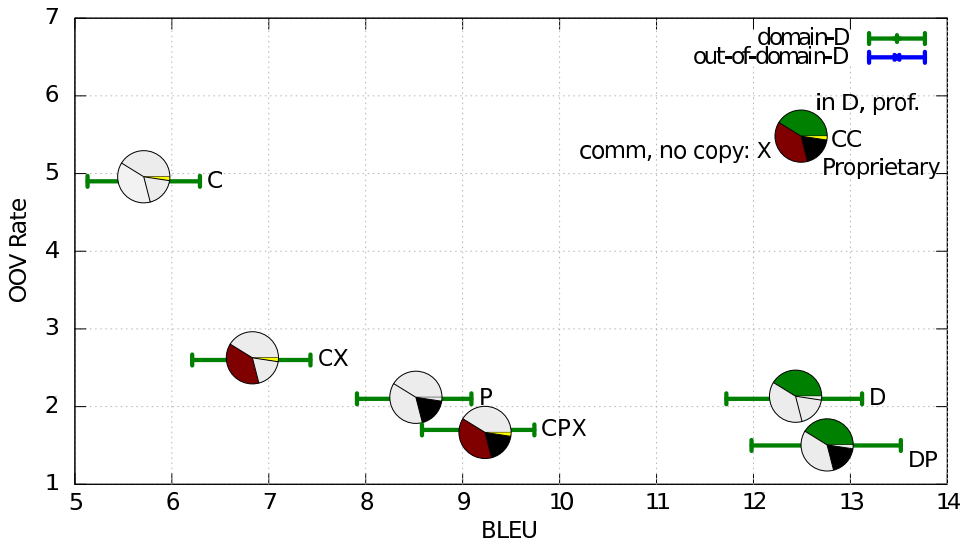
Vše mimo doménu



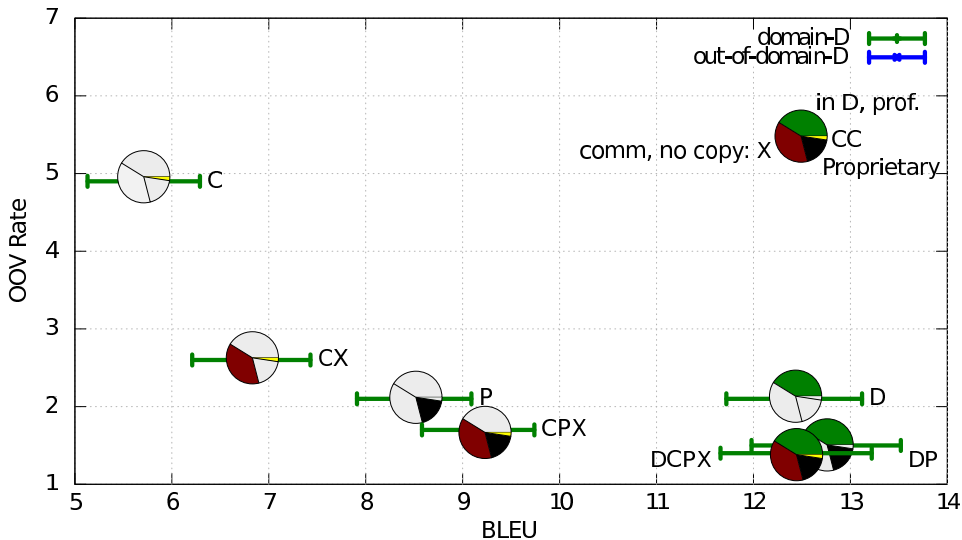
Obdobný objem v doméně: mnohem lepší



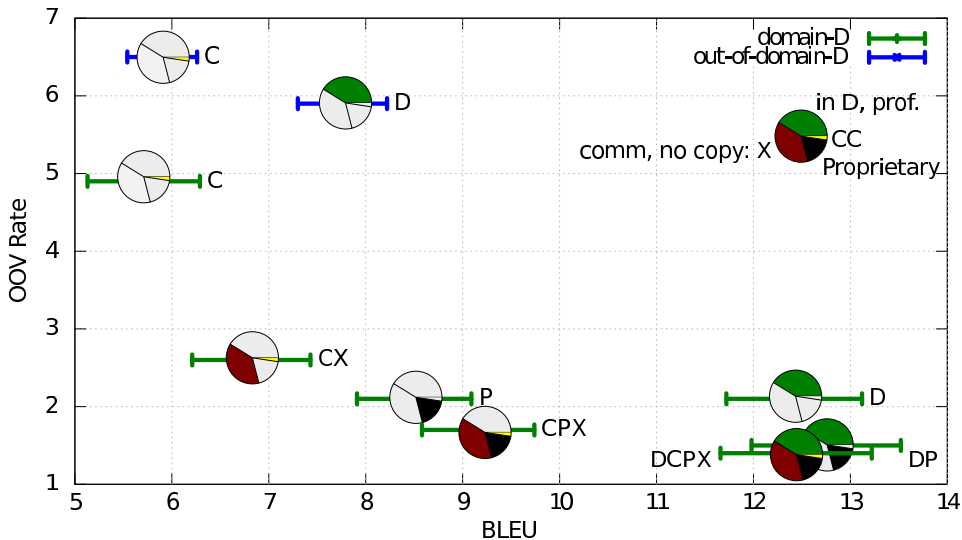
Dodatečná data zlepší pokrytí



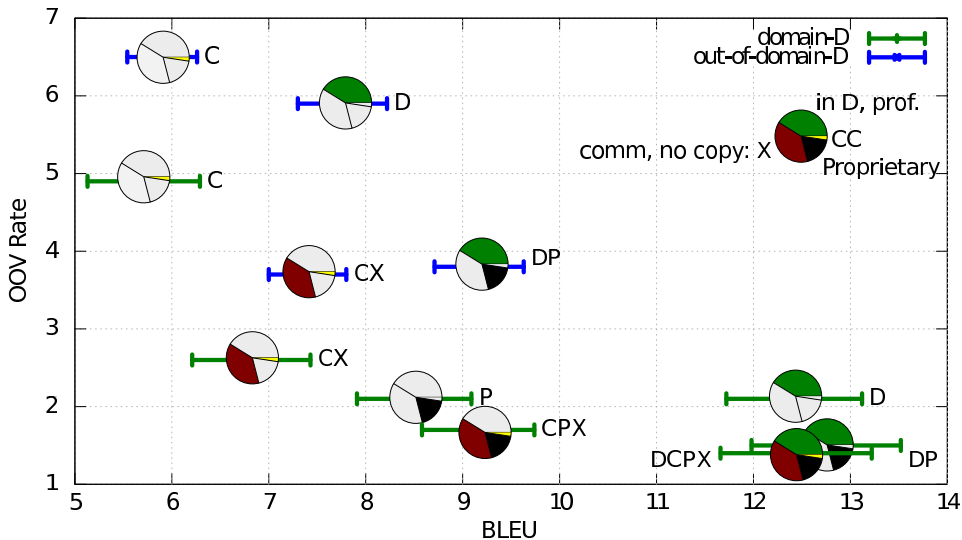
... ale mimo doménu mohou snížit kvalitu



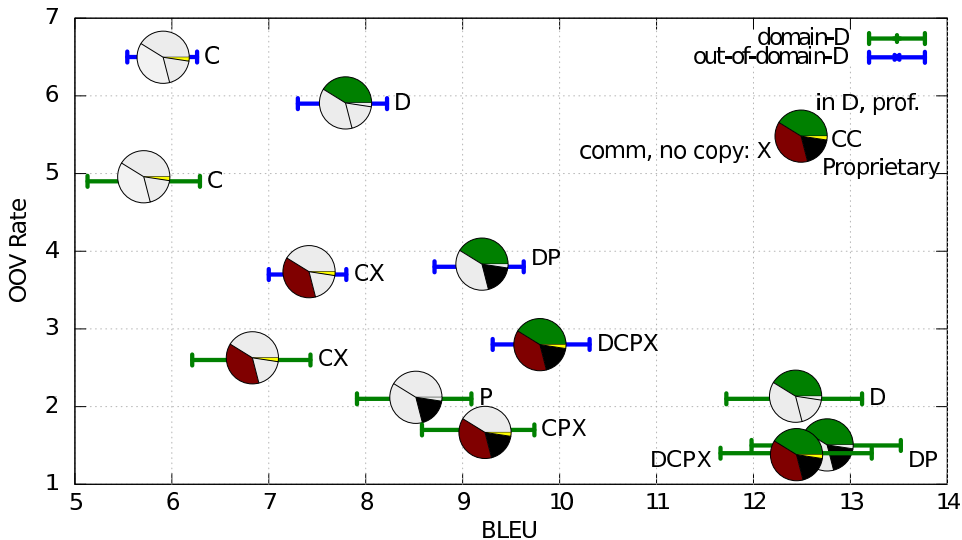
Mimo doménu: kvalita i pokrytí horší



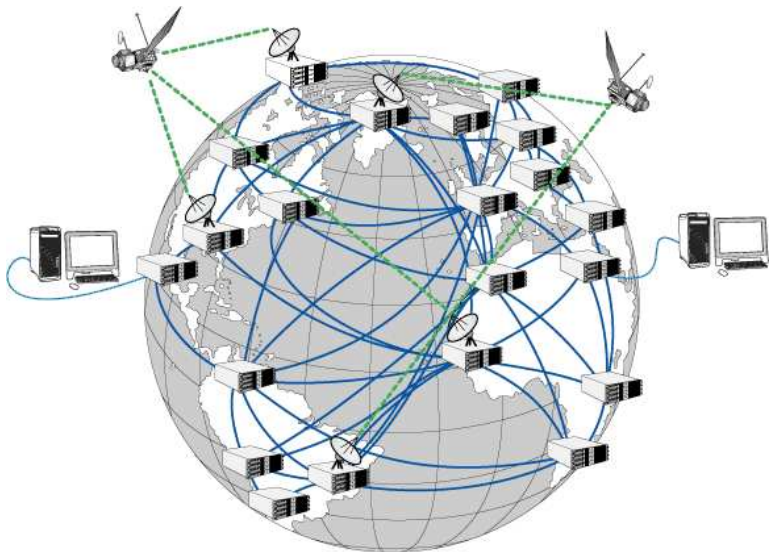
... víc dat zlepší pokrytí



... ale na moc kvalitu



Data raději z webu než z papíru



Paralelní webové stránky

File Edit View History Bookmarks Tije Tools Help

Travel to Norway - Official ... Visit Greece | Meetings and... Visit Greece | Συνεδριακός ...

www.visitgreece.gr/el/meetings_and_incentives

iGoogle Reading Groups Lit. Periodicals Recs GDT Οδηγία για το GDT Athens Google Map B2B EMTK ΜΑ/2008 Tréd User's Manual

Greece of art and science

Greece is a place of culture, the arts and sciences. Its tradition of contribution to global cultural and scientific communities, combined with its outstanding natural beauty and **excellent infrastructure**, has made it an ideal place in which to hold **conferences**. Over the last few years, Greece has more and more frequently welcomed people of letters, sciences and the arts, who have participated in symposia, conferences and exhibitions. **Athens International Airport 'Eleftherios Venizelos'**, one of the most modern airports in the world in operation since 2001, greatly boosted the organization of international conferences.

Η Ελλάδα των τεχνών και της επιστήμης

Η Ελλάδα αποτελεί έναν χώρο πολιτισμού, τέχνης και επιστημών. Η μακροίονη συμβολή της στο παγκόσμιο γίνεσθα, σε συνδυασμό με το μοναδικό φυσικό κάλλος και τις **άρτιες υποδομές**, την καθιστούν ιδανικό τόπο διεξαγωγής συνεδρίων. Τα τελευταία χρόνια, η ελληνική επικράτεια υποδέχεται όλο και συχνότερα ανθρώπους των γραμμάτων, των επιστημών και των τεχνών, οι οποίοι συμμετέχουν σε **συμπόσια, συνέδρια και εκθέσεις**. Ο **Διεθνής Αερολιμένας Αθηνών** «Ελευθέριος Βενιζέλος», ένα από τα πλέον σύγχρονα αεροδρόμια παγκοσμίως, ο οποίος λειτουργεί από το 2001, έδωσε μεγάλη ώθηση στη διοργάνωση διεθνών συνεδρίων.

Ο συνεδριακός τουρισμός είναι άκρως αλληλεπιδραστικός: απαιτεί, βέβαια, ένα υψηλό επίπεδο υπόβαθρο από τη χώρα υποδοχής, ταυτόχρονα όμως συμβάλλει ενεργά στην αναβάθμιση της συνολικής ποιότητας μιας περιοχής. Είναι λογικό, ένας χώρος ο οποίος προτιμάται για τη διεξαγωγή συνεδρίων, να μετέχει προνομιακά στο πολιτιστικό «προϊόν», μας και δίνει τη δυνατότητα σε κοινό, κατοίκους και επισκέπτες, να έρθουν σε επαφή με τα ανθρώπινα επιτεύγματα και τις καινοτομίες.

Conference tourism is extremely interdependent: it requires of course a high level of background support from the host country, and at the same time it can actively contribute to improving the overall standard of services in the region. It is logical that a country chosen as a conference location should be involved in the cultural 'product', giving the public, both residents and visitors, the chance to experience **human achievement and innovative thought**.

The Greece of the pre-Socratic philosophers, of the great poets, of Pheidias the sculptor

Η Ελλάδα των προσωκρατικών φιλοσόφων, των μεγάλων ποιητών, του Φειδία και του

Podstatná část dat nedostupná

OPENTEXT

The Deep Web

The Public Web

Only 4% of Web content (~8 billion pages) is available via search engines like Google

7.9
Zettabytes

The Deep Web

Approximately 96% of the digital universe is on Deep Web sites protected by passwords

Podstatná část dat nedostupná

SURFACE WEB

An iceberg floating in a blue ocean. The tip of the iceberg is above the water line and is labeled 'SURFACE WEB'. On the tip, the words 'Wikipedia', 'Google', and 'Bing' are written. The much larger part of the iceberg is submerged below the water line and is labeled 'DEEP WEB'. This submerged part is filled with various categories of information. At the very bottom of the iceberg, which is also submerged, is the label 'DARK WEB'.

Wikipedia

Google

Bing

DEEP WEB

Academic Information

Medical Records

Legal Documents

Scientific Reports

Subscription Information

Multilingual Databases

Conference Proceedings

Government Resources

Competitor Websites

Organization-specific
Repositories

DARK WEB

Illegal Information

Drug Trafficking sites

I dostupné texty nemusí být vytěžitelné

Národní referenční laboratoře SVS ČR

National Reference Laboratories of the SVA CR

Národní referenční laboratoř <i>National Reference Laboratory</i>	Sídlo <i>Address</i>
I. NRL pro potraviny a krmiva / NRL for Food and Feed	
1. NRL pro mléko a mléčné výrobky / <i>NRL for milk and milk products</i>	Státní veterinární ústav Praha <i>State Veterinary Institute Prague</i> Sídlištní 136/24 165 03 Praha 6 – Lysolaje
2. NRL pro salmonely / <i>NRL for Salmonellae</i>	Státní veterinární ústav Praha <i>State Veterinary Institute Prague</i> Sídlištní 136/24 165 03 Praha 6 – Lysolaje
3. NRL pro <i>Listeria monocytogenes</i> / <i>NRL for Listeria monocytogenes</i>	Státní veterinární ústav Jihlava <i>State Veterinary Institute in Jihlava</i> Rantířovská 93 586 05 Jihlava

A mnoho paralelních webů...

SVS - Chromium

SVS

eagri.cz/public/web/svs/portal

Veřejné zakázky Úřední deska SVS Tiskový servis Legislativa Kontakty E-podatelna Výsledky kontrol Česky English

Státní veterinární správa

eAGRI Hledaný výraz Podrobné hledání

Nejčastější dotazy

Státní veterinární správa Přihlásit

- o Státní veterinární správu
- Organizační útvary Státní veterinární správy
- Zaměstnání
- Základy a omezení
- Zdraví zvířat
- Pohoda zvířat - welfare
- Veřejné zdraví
- Laboratorní diagnostika
- Farmářské trhy
- Registrované subjekty
- Potraviny s chráněným názvem
- Formuláře ke stažení
- Legislativa
- Povinně zveřejňované informace dle zákona č. 106/1999 Sb.
- Dokumenty a publikace
- Odznaky
- Obchodování s veterinárním zbožím
- Nejčastější dotazy

Státní veterinární správa



Je orgánem státní správy v resortu zemědělství a je zřízena podle zákona č. 166/1999 Sb.

Jejím úkolem je především ochrana spotřebitelů před případnými zdravotně závadnými produkty živočišného původu, monitorování a udržování příznivé nálezové situace zvířat, veterinární ochrana státního území České republiky, ochrana pohody zvířat a ochrana před jejich týráním.

[Tisk stránky](#) [Poslat e-mailem](#)

zobrazit po 10 20 50

Prodejci vánočních kaprů budou mocí letos poprvé nahlašovat prodej také online

4.12.2015

Tisková zpráva – Vánoce se nezdářítečně blíží a již za necelé tři týdny se na ulicích a náměstích po celé České republice objeví tradiční kádě s živými kapry. Na to, aby prodej kaprů probíhal odpovídajícím způsobem, dohlédnou i letos úřední veterinární lékaři Státní veterinární správy (SVS). Prodejci živých ryb musí splňovat řadu povinností, mimo jiné alespoň sedm dní před zahájením prodeje nahlásit kde a kdy hodlají prodávat. Letos tak mohou prvně učinit také prostřednictvím online formuláře na internetových stránkách SVS.

Preventivní opatření v souvislosti s ptačí chřipkou

26.11.2015

Tisková zpráva – Ve Francii se poprvé od roku 2006 objevil u drůbeže vysoko nakažlivý druh aviární chřipky (subtypu H5N1). Výskyt ohniska na lhostěžné země ngvrdlív tamní úřad. Státní veterinární správa

Kontakt

Státní veterinární správa
Slezská 100/7, Praha 2, 12056
Adresa WWW: www.svs-cr.cz
Telefon: 227 010 142

[Kontaktní formulář](#)

Plášení zásilek potravin z EU



Důležité informace

- [Infekční anemie koní](#)
- [Africký mor prasat](#)
- [Výroční zpráva Státní veterinární správy za rok 2014](#)
- [Vysocepatogenní aviární chřipka](#)
- [Místa vstupu pro zvířata v zájmovém chovu](#)
- [Informace pro chovatele a obchodníky se skotem](#)
- [Dovoz potravin pro osobní](#)



... je paralelních jen zčásti.


State Veterinary Administration (SVS) - Chromium

State Veterinary Administration x

eagri.cz/public/web/en/svs/portal/

Press Service Contacts

Česky English

 **State Veterinary Administration**


eAGRI

Words to search

State Veterinary Administration

- › [About State veterinary administration](#)
- › [Animal welfare](#)
- › [Animal health](#)
- › [BIP in the Czech Republic](#)
- › [Registered subjects](#)
- › [Laboratory diagnostics](#)
- › [Published publications](#)
- › [Public Health](#)
- › [Trade with vet. commodities](#)

State Veterinary Administration



Is a public administration body under the Ministry of Agriculture of the Czech Republic and it was established according to the Veterinary Act No. 166/1999.

Its purpose is primarily the protection of consumers from products of animal origin likely to be harmful to human health, monitoring of animal health situation and maintaining it favourable, veterinary protection of the state territory of the Czech Republic, animal welfare and animal protection.

News

- [Results of Inspections in 2015](#)
Jan 1, 2015
- [TRAVELLING WITH PET ANIMALS](#)
- [Contamination of Food Chain by Residues - Situation 2013](#)
Jul 14, 2014

Postup

1. Najít dobré zdroje.
2. Získat surová data.
3. (Identifikovat a pojmenovat domény.)
4. Vyčistit.
 - ▶ Odstranit navigační prvky, reklamu, ...
 - ▶ Odstranit neparalelní části.
 - ▶ Odstranit jiné jazyky.
5. Zarovnat po dokumentech.
6. Zarovnat po větách.
7. Odstranit nekvalitně zarovnané části.
8. Zpřístupnit ostatním.
⇒ Paralelní korpus CzEng.

<http://ufal.mff.cuni.cz/czeng>

Šlo by to i lépe

- ▶ Můžete nám poradit dobré zdroje.
- ▶ Můžete nám pomoci s určením domény.
- ▶ Můžete nám přímo poskytnout své texty.

Šlo by to i lépe

- ▶ Můžete nám poradit dobré zdroje.
- ▶ Můžete nám pomoci s určením domény.
- ▶ Můžete nám přímo poskytnout své texty.



Shrnutí



YOUR DATA IS NEEDED!

Shrnutí

- ▶ Lingvistické analýzy potřebují anotovaná data v doméně.
- ▶ MT potřebuje **paralelní texty** v doméně i mimo ni.
- ▶ Čím víc, tím líp.
- ▶ Čím podobnější vstupům, tím lepší.

- ▶ Jazykové zdroje je třeba sdílet.
 - ▶ I průmysl data sdílí (např. TAUS).

- ▶ Zveřejněné překlady vytěžíme, netriviální úsilí.
 - ▶ Ideálně ne PDF, nejraději HTML s podobným URL.
- ▶ Mnohem lepší by byly primární zdroje
 - ▶ Vaše texty, překladové paměti.