

KWORDS

IKI words

Jazyk vstupního textu: Čeština ▼ Nápověda

Vstupní text (který chcete analyzovat):

- ▶ Vložte text
- ▶ Nahrajte textové soubory / Multi-analýza

Referenční korpus (úzus, s nímž chcete text porovnávat):

- ▶ Vyberte referenční korpus
- ▶ Vložte referenční text
- ▶ Nahrajte referenční textový soubor

Stop-list (slova, která mají být vyloučena z analýzy):

- zájmena
- předložky
- spojky
- čísla

Velikost písmen: ignorovat

Vítejte v **KWords**!

Je možné číst text z pohledu čtenáře z minulosti? Co bylo pro takového adresáta textu očekávané a co naopak překvapivé? Co se asi snažil autor zdůraznit v průběhu koncipování svého textu? Je vůbec možné zjistit, v čem se liší chápání stejného textu u různých adresátů? Lze nějakým způsobem změřit drobné a pozvolné společenské či kulturní změny, které se podepsaly na znění textu? Otázky podobné těmto bychom rádi zodpověděli s pomocí aplikace **KWords**. Tento nástroj je důležitou součástí společného projektu s názvem "[A Needle in a Haystack](#)", na němž se podílí Univerzita Karlova v Praze spolu s Brownovou Univerzitou v USA.

KWords slouží k identifikaci klíčových slov (tzv. keywords): to jsou slova (resp. slovní tvary), která jsou úzce spojena s hlavními tématy textu a s jeho žánrem. Klíčová slova tedy nejsou uživatelem předvybrána (jak by mohlo naznačovat jiné užití tohoto termínu např. ve vyhledávacích aplikacích), naopak, jsou to slova, která nástroj v textu odhaluje jako nejvíc prominentní. Tato aplikace umožňuje provádět corpus-driven výzkum slov, která jsou klíčová pro vyznění textu a následnou interpretaci, bez použití badatelovy intuice.

Výstupem **KWords** je seznam slov, jejichž frekvence je signifikantně vyšší v analyzovaném textu než v referenčním korpusu (např. v [SYN2015](#), což je vyvážený referenční korpus současné psané češtiny). Pokud je rozdíl ve frekvencích (po zohlednění odlišné velikosti textu a referenčního korpusu) statisticky významný (např. na základě testů χ^2 nebo log-likelihood), je následně vypočtena relevance rozdílu (tzv. effect size), podle níž jsou slova posuzována.

KWords

- slouží k identifikaci klíčových slov (tzv. keywords)
- Klíčová slova jsou slova (resp. slovní tvary), která jsou úzce spojena s hlavními tématy textu a s jeho žánrem.
- Klíčová slova tedy nejsou uživatelem předvybrána (jak by mohlo naznačovat jiné užití tohoto termínu např. ve vyhledávacích aplikacích), naopak, jsou to slova, která nástroj v textu odhaluje jako nejvíc prominentní.
- Tato aplikace umožňuje provádět corpus-driven výzkum slov, která jsou klíčová pro vyznění textu a následnou interpretaci, bez použití badatelovy intuice.

- Výstupem **KWords** je seznam slov, jejichž frekvence je signifikantně vyšší v analyzovaném textu než v referenčním korpusu např. v [SYN2015](#), což je vyvážený referenční korpus současné psané češtiny.

- **KWords** nadto ukazují i vztahy mezi klíčovými slovy.
- Pomocí funkce Keyword Links je možné sledovat, jak se klíčová slova (a témata, jež reprezentují) v analyzovaném textu spojují.
- Zároveň lze sledovat souvšskyty klíčových slov v grafu jejich distribuce napříč textem.
- Aplikace také umožňuje manuálně zkoumat konkordanční řádky, v nichž se klíčová slova vyskytují, a sledovat tak, do jakých kolokací vstupují v blízkém kontextu.



Analýza textu

UKÁZKA

Text

Klíčová slova

Distribuce

Keyword links

Konkordance

Než se pokusím odpovědět na otázku Kdo může v češtině **jit**?, podívejme se alespoň zevrubně na sémantické vymezení slovesa **jit**. Slovník spisovné češtiny (SSČ, 2000, s. 123) definuje sloveso **jit** jako "pohybovat se úmyslně (kroky) po vlastních nohou".

V prototypickém významu sloveso **jit** označuje fyzický **pohyb** z jednoho místa na jiné, tj. z nějakého východiska do cíle **pohybu**. Východisko i cíl mají povahu fyzického prostoru, pohybujícím subjektem je prototypicky člověk, **pohyb** probíhá po pevném povrchu a je složen z dílčích pravidelně se opakujících jednotek, tj. kroků.

Cílem tohoto textu však není popsat metaforickou extenzi slovesa **jit**, zaměříme se pouze na možnost obsazení jeho subjektivního participantu se selektivním sémantickým rysem +animal, tj. na to, o kterých zvířatech můžeme říci, že jdou. Protože sloveso **jit** patří mezi nejfrekventovanější česká slovesa (Bartoň, 2009) a protože "svět **zvířat** je vždy nahlížen z lidské perspektivy (a) oba světy jsou vzájemně propojené a neoddelitelné" (Šťastná, 2019, s. 481), lze předpokládat, že sloveso **jit** zachycující lidský **pohyb** pomocí nohou, resp. kroků, přenášíme antropomorficky i na **zvířata** a popis jejich **pohybu**. Lokomoce živočichů se rozlišuje podle prostředí, v němž se **pohyb** uskutečňuje, běžně se rozlišuje "chůze (běh), plavání nebo let".

Kompatibilita **pohybu** vyjádřeného pomocí slovesa **jit** se tedy nevztahuje na celou skupinu **zvířat** absolutně. V učebnicích češtiny pro cizince ani ve slovnících neexistují instrukce, pro která **zvířata** je možné sloveso **jit** použít. V roce 2019 jsem proto realizovala malou průzkumnou sondu mezi českými studenty programu Filologie na FP TUL a nerodilými mluvčími češtiny studujícími v programu Bohemistika pro cizince na FF UK. Rodilí i nerodilí **mluvčí** dostali stejný úkol, jehož zadání znělo "Označte živočichy, jejichž **pohyb** můžete označit pomocí slovesa **JIT**, tj. např. zebra jde. Pokud **pohyb** zvířete nemůžete označit slovesem **jit**, uveďte sloveso vystihující **pohyb** zvířete." **Mluvčí** vybírali z následujících **zvířat** řazených v uvedeném pořadí: morče, pes, kůň, holub, prase, pavouk, kočka, kráva, krokodýl, želva, moucha, slepice, slon, králík, veverka. Z množiny vybraných **zvířat** se všechna pohybují po pevném povrchu, všechna mají nohy, byť v různém počtu.

Vybraná **zvířata** se záměrně liší svou velikostí vůči člověku, který jejich **pohyb** označuje a je kognitivně nucen srovnávat svůj **pohyb** chůze s pohybem, který vykonávají **zvířata**. Dalším kritériem výběru byla rychlost **pohybu** a délka končetin **zvířat** vzhledem k jejich tělu. Tato dvě kritéria jsou důležitá pro vnímání kroku a pravidelnosti střídání tohoto **pohybu**. Ve skupině jsou proto zařazena **zvířata** poměrově výrazně větší i menší než člověk; s dvěma, čtyřma, šesti a osmi nohama, jejichž velikost je k poměru těl **zvířat** výrazně větší i výrazně menší než tělo zvířete.

S mluvčími obou skupin jsem následně diskutovala o jejich výběru sloves a o zvířatech, u nichž s užitím slovesa **jit** váhali, a o tom, co jejich váhání způsobovalo. Rodilí **mluvčí** měli následně zvažovat kritéria kompatibility, resp. nekompatibility slovesa **jit** se zvolenými zvířaty.

Všichni nerodilí **mluvčí** bez váhání spojili všechna uvedená **zvířata** se slovesem **jit**. Při první konfrontaci nebyli většinou schopni uvádět, zda nějaké antropomorfické rysy zvířete a způsob **pohybu** souvisejí s výběrem slovesa **jit**. Při diskusi byli schopni tento **pohyb** konfrontovat pouze s pohybem létání a plavání, příp. plazení. Rodilí **mluvčí** naopak hned při prvním čtení vykazovali v přiřazování **zvířat** k pojmenování jejich **pohybu** rozdělení do tří typů skupin **zvířat** vzhledem k uplatnění parametru kompatibility se slovesem **jit**.

První skupinou **zvířat** jsou **zvířata** jednoznačně spojitelná se slovesem **jit**. Mezi tato **zvířata** se řadí pes, kůň, kráva, slon. Z diskuse se studenty vyplývá, že **pohyb** těchto **zvířat** označují slovesem **jit**, protože zřetelně rozeznávají střídavý kmit nohou, který "není moc rychlý" a je dobře patrný, protože **zvířata** jsou dostatečně velká a rozměry jejich nohou jsou při percepci dobře rozeznatelné vzhledem k jejich tělu.

Druhou skupinou jsou **zvířata** pro rodilé **mluvčí** se slovesem **jit** jednoznačně nespojitelná. Tato skupina zahrnuje **zvířata** pavouk, krokodýl, želva, moucha, králík, veverka. Tato **zvířata** jsou buď moc malá a **pohyb** jejich nohou je tak rychlý, že je těžko zrakově zaznamenatelný, resp. spojitelný s pravidelným střídáním kroku (pavouk, moucha), nebo je naopak příliš pomalý a střídavý **pohyb** všech končetin není vzhledem k velikosti těla dobře sledovatelný (želva, krokodýl), nebo je příliš rychlý a nevykazuje pravidelný rytmus střídání končetin (veverka, králík). Pro tato **zvířata** **mluvčí** raději volí spojení se slovesy **pohybu** běžet, lézt, skákat, hopkat, atp.

Třetí skupinu pak představují **zvířata**, u nichž se rodilí **mluvčí** na kompatibilitě se slovesem **jit** neshodují. Mezi tato **zvířata** patří morče, kočka, holub, prase, slepice. U těchto **zvířat** se **mluvčí** různí v interpretaci **pohybu**. Někteří rodilí **mluvčí** označují **pohyb** těchto **zvířat** slovesem **jit**, **pohyb** nohou vnímají jako pravidelný a dobře vnímatelný vzhledem k velikosti těla **zvířat**. Druhá část rodilých mluvčích však užití slovesa **jit** pro **pohyb** daných **zvířat** považuje za nepatřičné. Roli v této interpretaci hraje rychlost **pohybu** nohou, u morčete, kočky a prasete někteří rodilí **mluvčí** vnímají **pohyb** jako příliš rychlý a označují jej už jako běh. U slepice a holuba, ačkoliv se pohybují po dvou nohou tak jako člověk, naopak vnímají **pohyb** jako příliš pomalý, ale také poukazují na to, že se jeví jako příliš důkladný, kvůli této důkladnosti jej označují raději pomocí slovesa kráčet. U kočky je pro některé **mluvčí** její **pohyb** buď příliš rychlý, tj. kočka běží, nebo příliš pomalý kačka se plíží/našlapuje, neodpovídá však chůzi.

Závěrem lze shrnout, že **pohyb** popisovaný jako chůze a vyjadřovaný v češtině slovesem **jit**, není v češtině přisuzován výlučně člověku, ale přenášíme jej i na **zvířata**. **Zvířata**, která se pohybují po zemi pomocí nohou, pak lze vnímat jako škálu umístěnou mezi kladným bodem, který zahrnuje **zvířata**, která kolokaci se slovesem **jit** jednoznačně vytvářejí, a záporným bodem, který zahrnuje **zvířata**, která kolokaci se slovesem **jit** jednoznačně nevytvářejí. Mezi těmito body jsou umístěna **zvířata**, u nichž se na označení **pohybu** pomocí slovesa **jit** rodilí **mluvčí** neshodnou, a to zvl. z důvodu rychlosti **pohybu** nebo poměru velikosti nohou vůči tělu zvířete.

KLÍČOVÁ SLOVA

Text

Klíčová slova

Distribuce

Keyword links

Konkordance

Shrnutí

	Vstupní text	Referenční korpus: vlastní
Tokeny	1128	606
Typy	446	317

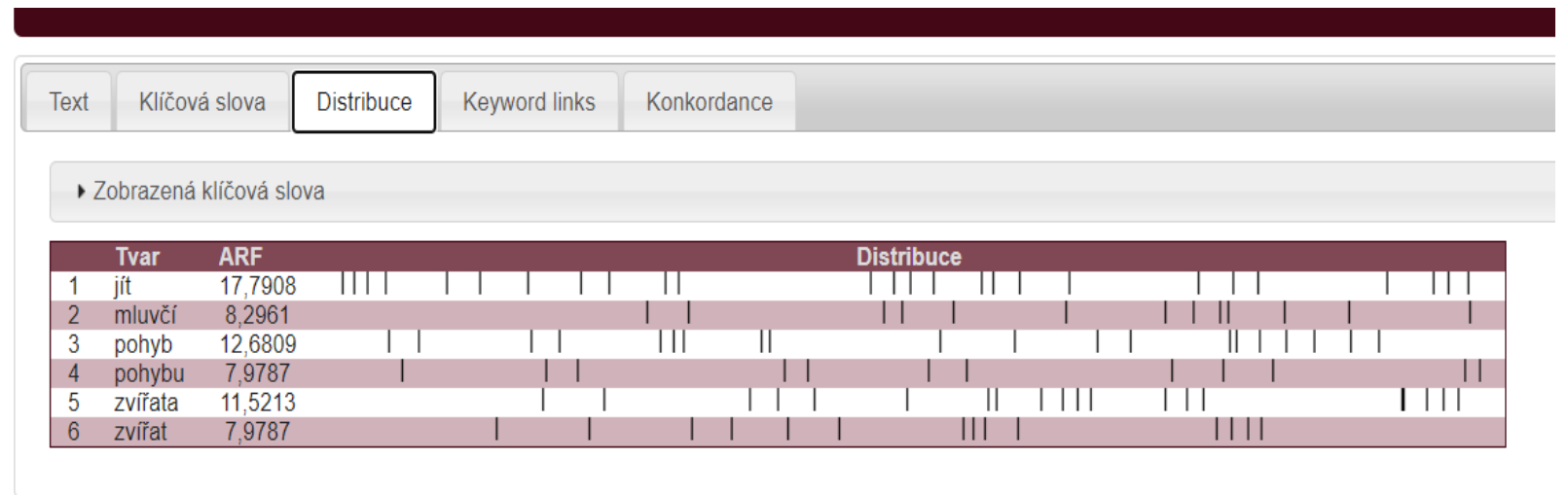
Tematická koncentrace

Pořadí	Tvar	FQ
3	jít	26
4	zvířata	20
5	pohyb	20
6	zvířat	14
8	mluvčí	13
11	pohybu	12

Klíčová slova

	Tvar	LL	DIN	Fq(text)	Fq(ref)	TC
1	jít	22.000	100.0000	26	0	+
2	mluvčí	11.000	100.0000	13	0	+
3	pohyb	17.000	100.0000	20	0	+
4	pohybu	10.000	100.0000	12	0	+
5	zvířata	5.000	56.0000	20	3	+
6	zvířat	1.000	30.0000	14	4	+

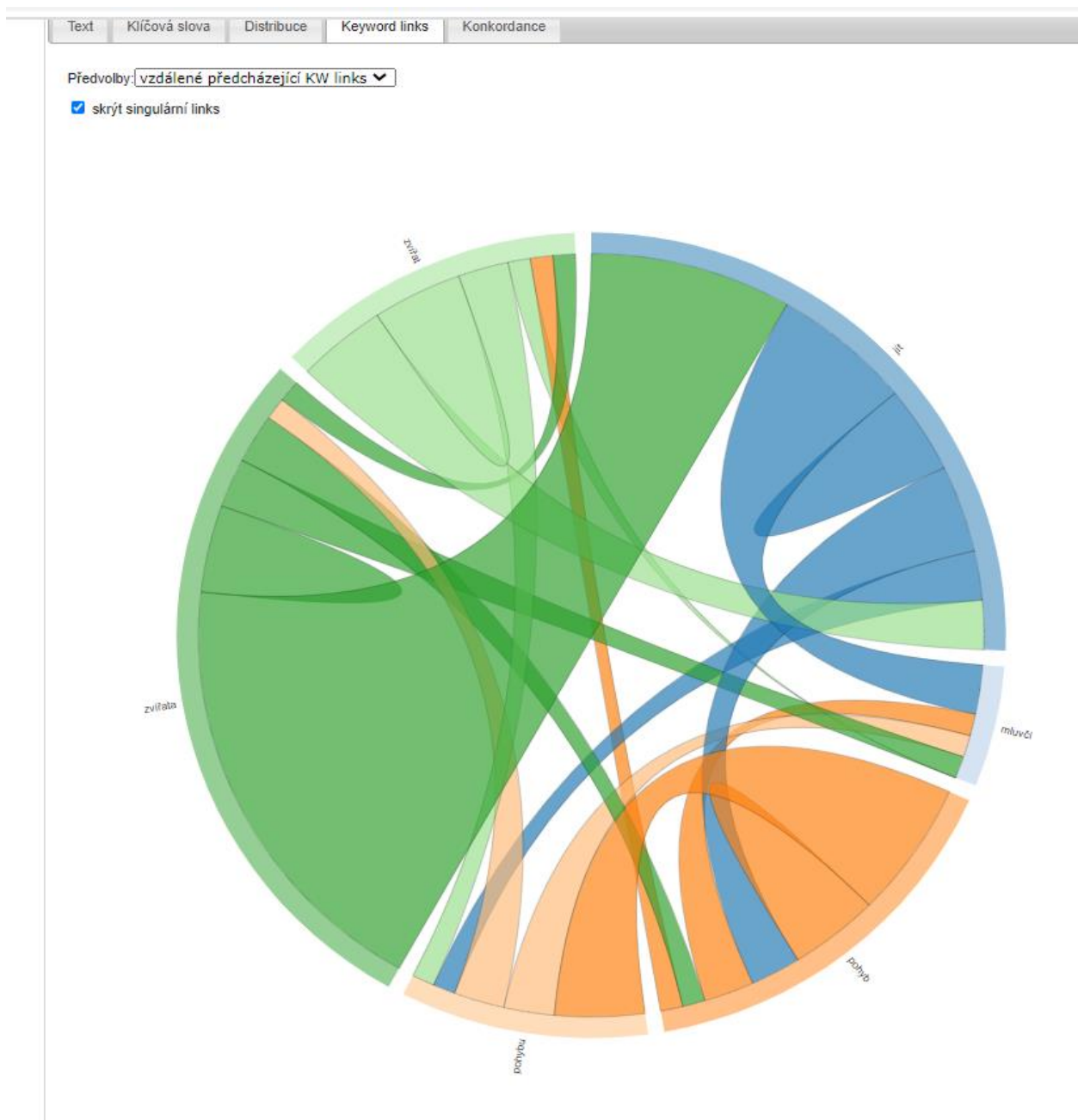
DISTRIBUCE
Ize sledovat
souvýskyty
klíčových slov
v grafu jejich
distribuce
napříč textem

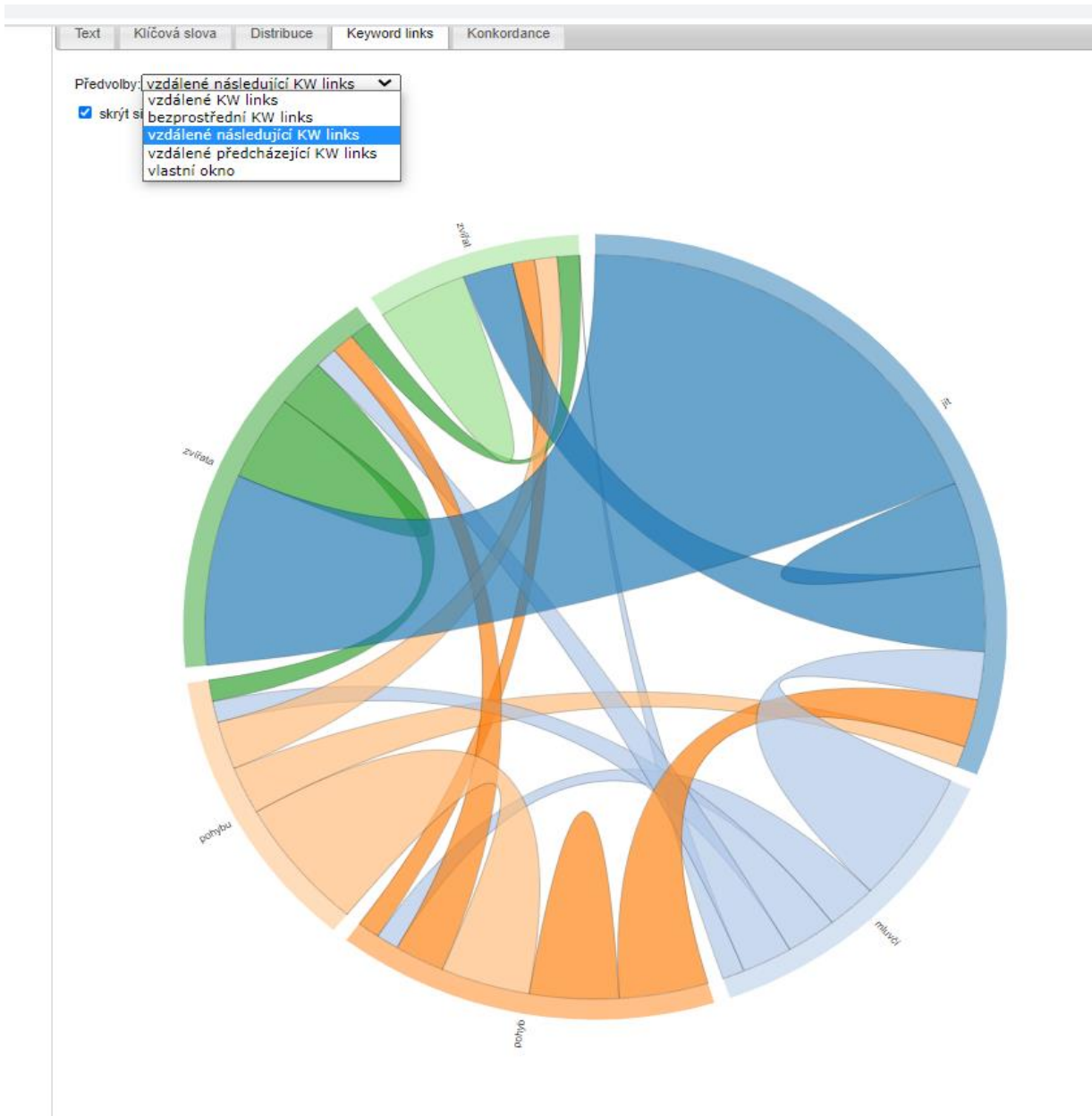


KEYWORD LINKS

ukazují vztahy mezi klíčovými slovy.

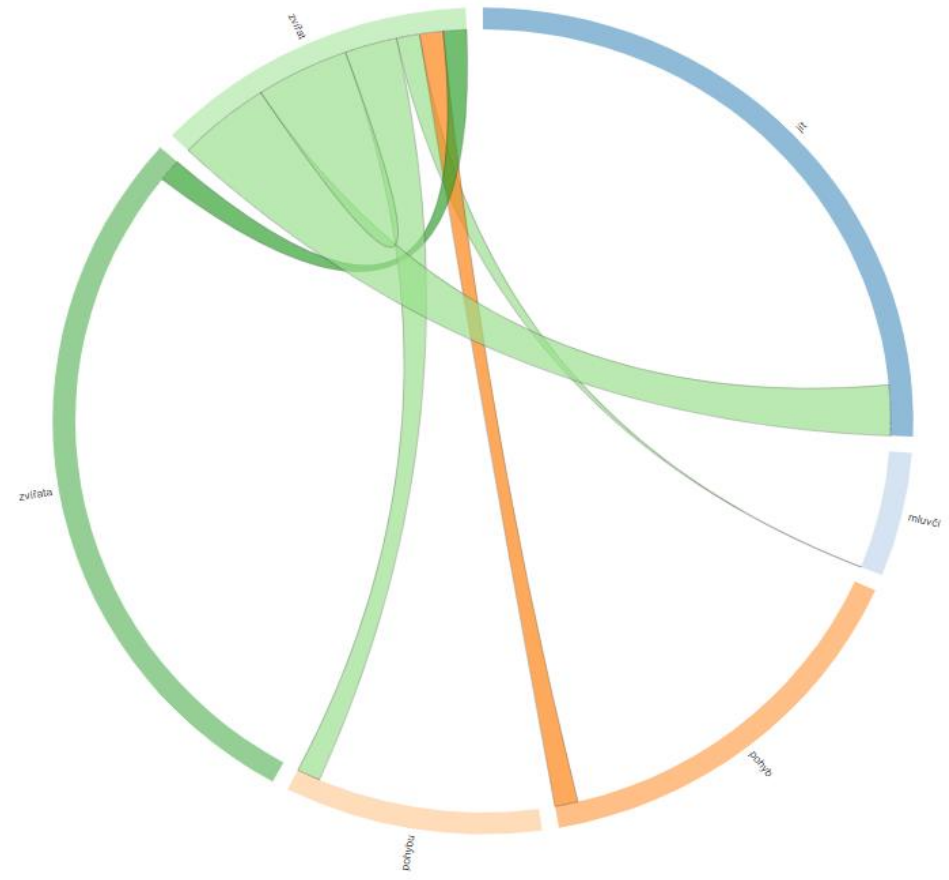
je možné sledovat, jak se klíčová slova (a témata, jež reprezentují) v analyzovaném textu spojují.





Předvolby: vzdálené předcházející KW links ▾

skrýt singulární links



umožňuje
manuálně
zkoumat
konkordanční
řádky, v nichž se
klíčová slova
vyskytují, a
sledovat tak, do
jakých kolokací
vstupují v
blízkém
kontextu.

Analyzovat

Text Klíčová slova Distribuce Keyword links Konkordance

nejfrekventovanější slova na předcházející pozici	klíčové slovo	nejfrekventovanější slova na následující pozici
slovesem (11) • slovesa (9) • sloveso (5) • češtině (1) rodilí (6) • nerodilí (2) • zvířata (1) • rodilé (1) , (2) • vnímají (2) • že (2) • svůj (1) rychlost (2) • jejich (2) • interpretaci (1) • cíle (1) těchto (3) • končetin (1) • skupin (1) • vybraných (1) tato (4) • zahrnuje (3) • na (2) • jsou (2)	jít (26) mluvčí (13) pohyb (20) pohybu (12) zvířat (14) zvířata (20)	, (6) • . (5) • jednoznačně (3) • se (2) její (1) • vybírali (1) • na (1) • naopak (1) zvířete (2) • těchto (2) • jako (2) • daných (1) . (4) • nebo (1) • rozdělení (1) • pomocí (1) vzhledem (2) • se (2) • absolutně (1) • výrazně (1) , (5) • se (3) • . (2) • jsou (2)