دۆخی ئێستای تەکنۆلۆژیای زمان بۆ کوردی
چ کراوە و چ بکرێ؟

سینا ئەحمەدی

https://sinaahmadi.github.io

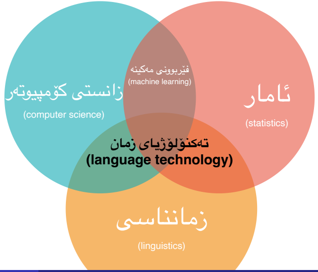
ناوەرۆک

1. تەکنۆلۆژیای زمان
2. زمانی کوردی
3. تەکنۆلۆژیای زمان بۆ کوردی
4. دەرنجامەکان

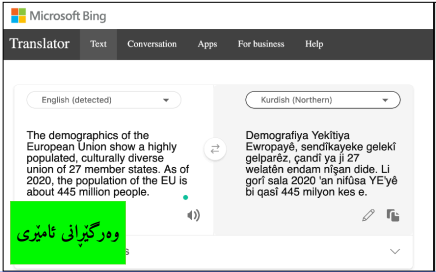
تەکنۆلۆژیای زمان

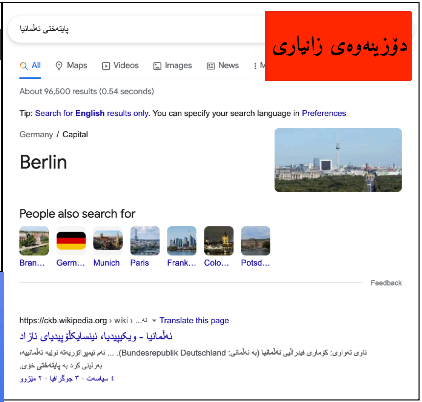
تەکنۆلۆژیای زمان: تێگەیشتنی « کۆمپیوتەری » ی زمان

- **زمانەوانیی کۆمپیوتەری** (computational linguistics – CL): لێکدانەوەی زمانەکان بە یارمەتیی تەکنیکە کۆمپیوتەرییەکان.
  ⇐ بۆچوونی زانستیی زمانەوانەکان بۆ تێگەیشتن لە زمان.
- **پرۆسەس کردنی زمانی سروشتی** (natural language processing – NLP): سازکردنی ئامێر، ئەلگۆریتم و سەرچاوەکان بۆ چارەسەرکردنی پرۆسەس کردنی زمان.
  ⇐ بۆچوونی ئەندازیاری بۆ چارەسەرکردنی کێشەکان و خوڵقاندنی تەکنۆلۆژیا.
- **زمانەوانیی کۆمپیوتەری، پرۆسەس کردنی زمانی سروشتی و تەکنۆلۆژیای زمان** (language technology) زۆر جار وەکوو یەک بەکاردەبرێن.



تەکنۆلۆژیای زمان: هەندێک لە ئەرکەکانی تەکنۆلۆژیا





- دەق
- بە شێوەیەکی نەریتی، زۆرتر دەقی هەواڵەکان بەکاردێن، وەک لە ڕۆژنامەکان و ماڵپەڕە هەواڵنێرییەکان؛ زۆربەیان خاوێن، فۆرماڵ و ستانداردن.
  - کارە نوێیەکان هەمەچەشنترن و دەقی زۆر بوار، بابەت و بنزاراوەی دیکە لە خۆ دەگرن، وەک بلاگەکان، تویتەر یان زمانی ئاماژە.
- دەنگ
  - ئاخاوتن ئالۆزترە، ناڕێککەوتن زۆرترە و زمانی نافەرمیشی تێدایە.
  - بۆ ناسینەوەی ئۆتوماتیکی ئاخاوتن و گێڕانەوەی دەق بۆ ئاخاوتن بەکاردێن.
- سەرچاوەی زۆربەی پێدراوەکان وێبە و سەرچاوە کراوەیە (open-source).

تەکنۆلۆژیای زمان: زانیارییەکان بۆ زمان

- زانیاری لە سەر فۆنتیک و فۆنۆلۆژی
- زانیاری لە سەر مۆرفۆلۆژیی و جیاوازیی نێوان شێوەزارەکان
- زانیاری لە سەر سینتاکس و پێکهاتەکانی ڕستە
- زانیاری لە سەر سێمانتیک، پراگماتیک و دیسکۆرس
- **زانیاری لە سەر دنیا**

گەورەترین شاری کوردی ڕۆژهەڵات چ شارێکە؟

تەکنۆلۆژیای زمان: زمانەکان لە دنیادا

- پتر لە 7000 زمان لە دنیا قسەیان پێدەکرێ [1]
- بەشێکی زۆری ئەم زمانانە، سەرچاوەی زمان و زمانەوانییان بۆ کەمە [2]
- وێڕای پێشکەوتنەکان لە زیرەکیی دەستکرددا، زۆربەی زمانەکان پێویستییان بە پسپۆڕایەتیی زمانەوانییە.

زمانە سەرچاوە دەوڵەمەندەکان

سەدان ملیۆن بەڵگەی سەرهێڵ و پێدراوە و زانیاری
وەک ئینگلیسی

زمانە مامناوەندییەکان

ملیۆنان بەڵگەی سەرهێڵ و پێدراوە و
زانیاری زۆربەی زمانە ئەورووپییەکان

زمانە هەژارەکان

بێ یان کەم/سەرچاوە
زۆربەی زمانەکانی دنیا

تەکنۆلۆژیای زمان: مەترسیی لەناوچوونی زمانەکان

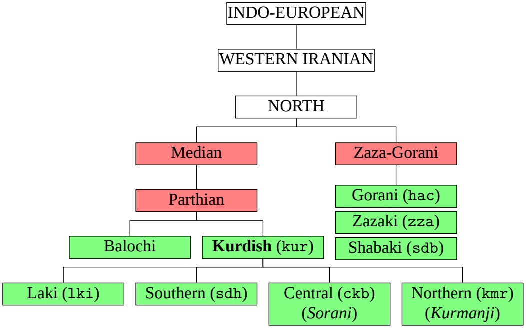
بەڕای دەیڤید کریستاڵ [3] زمانێک لە مەترسیی لەناوچوون پێش دەکەوێت ئەگەر پەیڤەرەکانی...

- پریستیژی خۆیان لە کۆمەڵگای زاڵدا زۆر بکەنەوە.
- دەوڵەمەندییان لە حاندی کۆمەڵگای زاڵدا زۆر بکەنەوە.
- هێزی دەسەڵاتی خۆیان لە چاوی کۆمەڵگەی زاڵدا زۆر بکەنەوە.
- لە سیستەمی فێرکاریدا بوونێکی بەهێزی هەبێ.
- بتوانن بە زمانەکەیان بنووسن.
- **لە تەکنۆلۆژیی ئلکترۆنیک کەڵک وەرگرن.**

زمانی کوردی

زمانی کوردی

- زمانێکی هیندوئەورووپییە
- زۆرتر لە 30 ملیۆن ئاخێوەری هەیە
- زۆری بنزاراوە و شێوەزار هەیە (شێوەزار یان زمان؟)
- چەند ئەلفوبێی هەیە کە تێیاندا ئی عەرەبی و لاتین باوتر بەکاردەهێندرێن
- مێژوویەکی درێژخایەنتری لە گێڕانەوەی زارەکیدا هەبووە هەتا نووسین
- زمانێکە کە سەرچاوەی بۆ کەمن.



سەرچاوەی وێنە: [4]

سەرچاوەی وێنە: https://www.britannica.com/topic/Kurd

زمانی کوردی: هەندێک لە کێشەکان

- یەکلانەبوونەوەی دۆخی زمانی فەرمی و ڕێنووسی ستاندارد
- بەکار(نە)هێنانی چەند ئەلفوبێ، ڕێنووس و شێوەی نووسین
  - « زیاتر وبةختةوةر بزیت کاکة گولةکةم »
  - « وولات » یان « وڵات »؟ « درك » یان « درک »؟
  - *di sala 2020'an | 2020-an | 2020an de*
  - *hêv**i**ya | hêv**î**ya | hêvî ya*
  - ٠١٢٣٤٥٦٧٨٩ | ۰۱۲۳۴۵۶۷۸۹ | 0123456789
- گەرچی دەگوترێ ئەلفوبێیەکانی کوردی فۆنێمیکن، واتە هەر دەنگەو پیتێکی هەیە، هەمیشە وا نییە!
  - پیتە دوولاکان: « ی » بۆ « y/î » و « و » بۆ « w/u »
  - جیاوازییەکان لە هەندێک لە ئەلفوبێیەکان وەک « l »، « ll » یان « ł » بۆ [ɫ]
  - پیتە ونەکان وەک بزرۆکە (i) لە ئەلفوبێ عەرەبی وەک « بزر » (bizr/bizir)

[lki-ar]
ئیمرووژ پەڵاونێ مەرەکە گرتوتێ خەڵکیژ بێ هوول ئەژ کورونا دەوران گرتو

[sdh-ar]
فەڵسەفە وەرجە سۆقرات، چاودێر زانستەیل سرووشتی بۆیە و کارێگەو کردار، باوەڕ، دین و ئاین خەڵک نیاشتێیە

[kmr-ar]
وەزارەتا ئەوقافێ وکاروبارێن ئاینینی ل هەرێما کوردستانێ ل دۆر بێهنڤەدانەکا فەرمی ب هەلکەفتەکا ئاینینی رۆهنکرنەك دەرکر

[ckb-ar]
لە ڕاستیدا ئەم کارەکتێرانە سەر بە کۆمەڵگای سوننەتیی کوردستان و جیلەکانی ڕابردوون

[kmr-latn]
Ji ber barîna berfê li bajarê Wan û navçeya Tetwan a Bedlîsê dîmenên ciwan derketin holê.

[ckb-latn]
Bergirî lem bwareda her le yekemîn rojekanî damezrandinî komarî Turkyawe hate gořê.

تەکنۆلۆژیای زمان بۆ کوردی

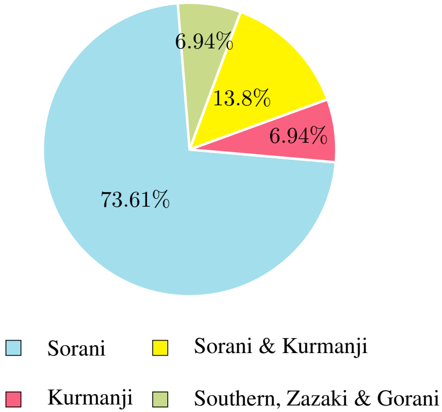
تەکنۆلۆژیای زمان بۆ کوردی: پرسیار

سەردانی ئێرە بکە و وڵامی پرسیارەکان دەوە:

https://kahoot.it

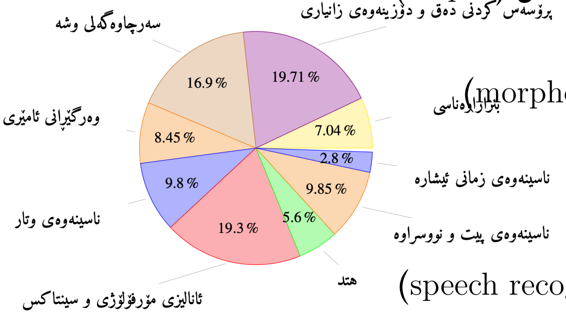
تەکنۆلۆژیای زمان بۆ کوردی: هەندێک زانیاری

- توێژینەوەی تاک و تەراک
  ⇐ کەمتر لە 80 وتاری زانستی بڵاوکراوەتەوە (ڕێک 72 هەتا حوزەیرانی 2022) [5]
- یەکەم کارەکان لەم بوارەدا دەگەڕێنەوە بۆ ساڵی 2009
  بێجگە لە بەرهەمێکی دوکتۆر شێرکۆ بابان لە ساڵی 1995دا [6]!
- ڕێژەیەکی کەمیان سەرچاوەکانیان کراوەیە.
- هەموو بنزاراوەکان بە یەک شێوە تیشکیان لە سەر نەخراوە.
- لە دواوەین بەڵام بەرەبەرە ئاگادارییی گشتی زۆرتر بووەتەوە.



سەرچاوەی وێنە: [5]*

- ناشارەزاییی کوردیئاخێوان لە ناسینی زمانەکەیان
  ⇐ بەکارنەهێنانی لە سەر تۆڕە کۆمەڵایەتییەکان
  ⇐ **پێدراوەی کەم**
- کێشەکانی ئەلفوبێ و ڕێنووس
- دەوڵەمەندیی کوردی و بنزاراوەکانی
  « گوڵەکانت زۆر جوانن » / « گوڵەکانت فرە زەریفن » / « گورەکانت زۆر جوانن» .
- کەم/نەبوونی پێدراوە بۆ سەرجەم زاراوەکان بەیەک شێوە
- دەوڵەمەندیی مۆرفۆلۆژیی کوردی
  - داشماننەگرتبووبان ← دا-ش=مان-نە-گرت-بوو-با-ن
  - تێشیاندایە ← تێ-ش=یان=دا-یە



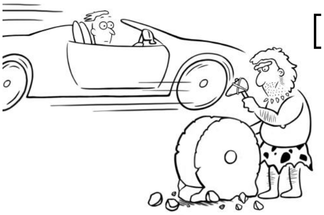
سەرچاوەی وێنە: [5]*

- کۆکردنەوەی سەرچاوەی زمان وەک کۆرپسەکان (corpus)
- هەڵەگریی ڕێنووس (spell checking)
- ئانالیزی مۆرفۆلۆژی (morphological analysis)
- لێکدانەوەی وشە و دیتنەوەی سنووری ڕستە و وشەکان (tokenization)
- گێڕانەوەی ئەلفوبێ (transliteration)
- ڕێکخستنی سیستەمی ئاخاوتن (speech recognition)
- وەرگێڕانی ئامێری (machine translation)
- هتد.

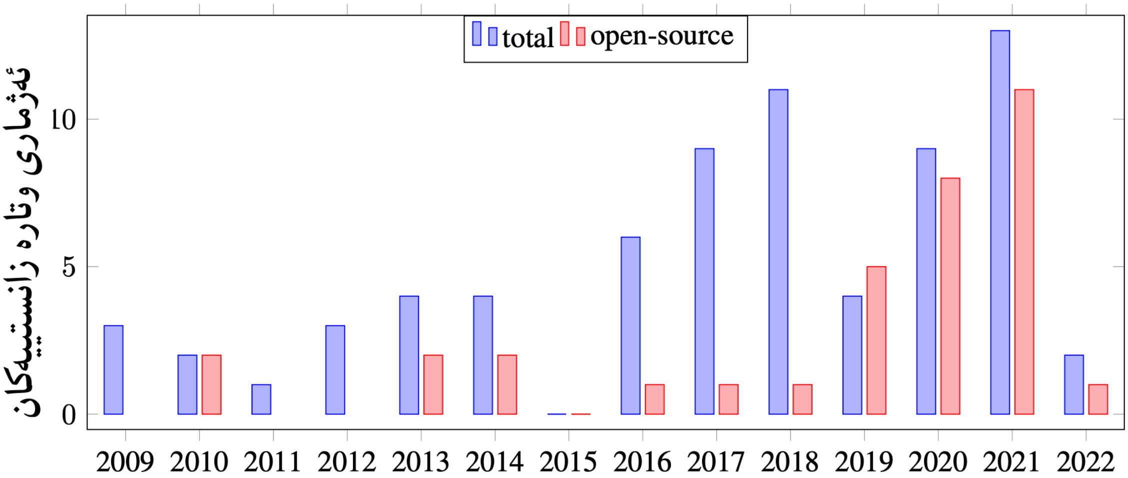
کراوەیی سەرچاوە (open-source)

نەرمەکاڵا یان بەرهەمی سەرچاوەکراوە بە شێوەیەکی کراوە بۆ هەمووان بەردەستە.

- هەندێک لە پڕۆژەکان لە سەر باسێک کە پێشتر باسی کراوە ورد دەبنەوە بە بێ چارەسەری پراکتیکی
  ستێم دیتنەوە (stemming) لانی کەم **پێنج** جاران باسی کراوە [5]
- هەندێک لە پڕۆژەکان بەزەحمەت لە پڕۆژەی گەورەدا بە کار بێن
  فەرهەنگێکی گەورەی مۆرفۆلۆژی بۆ زمانی کوردی لە چوارچێوەی Alexina [7]
- هەندێکیشیان بە شێوەیەکی نامۆمەڵ بڵاو کراونەتەوە
  بۆ نموونە ئامێری وەرگۆڕانی دەقی (transliteration) [8]
- **کەوا، زمانی کوردی هێشتا هەر هەژارە!**



سەرچاوەی وێنە: [5]*


ژمارەی وتارە زانستییەکان
total
open-source
0
5
10
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022

تەکنۆلۆژیای زمان بۆ کوردی: هەندێک پرۆژە

- هەندێک پرۆژەی زۆر باش وەک
  - ویکیفەرهەنگ: https://ku.wiktionary.org
  - ویکیپیدیای کوردی: https://ckb.wikipedia.org
  - ڤەژینبوکس https://books.vejin.net/ و ڤەژینلێکس https://lex.vejin.net/
  - ئامێری پرۆسەس کردنی زمانی کوردی (KLPT) https://github.com/sinaahmadi/klpt

- گرینگایەتیی کوردی لە چاو گووگڵ، ئامازۆن، فەیسبۆک و ئەپڵ (گافە)دا
  - کوردی لە سەر وەرگێڕی گووگڵ (https://translate.google.com/)
  - کوردی لە سەر وەرگێڕی مایکرۆسۆفت (https://www.bing.com/translator)

- ئامێرێکی بنەڕەتی بەڵام دەکرێ بەئاسانی درێژە بدرێتێ
- ئامانجی سەرەکیی پرۆسەس کردنی زمانی کوردی بە هەموو بنزاراوەکان و ئەلفوبێیەکانییەتی
- هیچ کتێبخانەیەکی دەرەکی تێیدا بەکارنەهێنراوە
- ئیلهام لە کارایی ئامرازە پەیوەندیدارەکان وەرگیراوە وەک NLTK و spaCy
- کراوەیە!

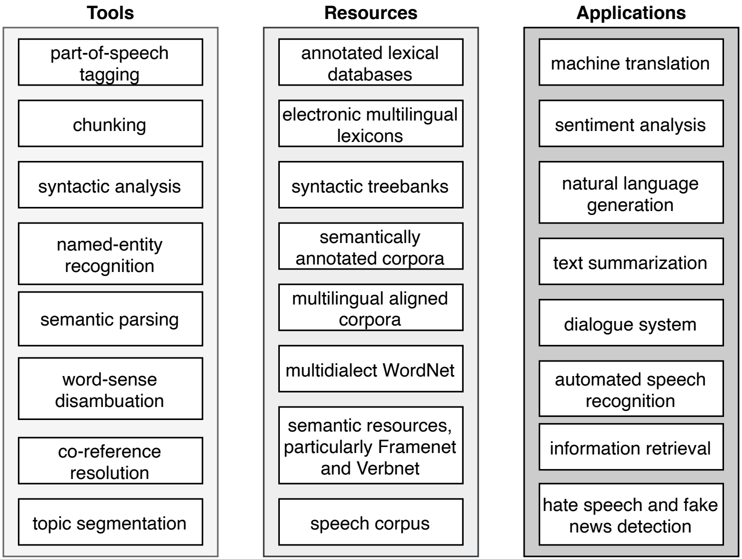
https://github.com/sinaahmadi/klpt

تەکنۆلۆژیای زمان بۆ کوردی: ئامێری پرۆسەس کردنی زمانی کوردی

- بۆ ئێستا بۆ کورمانجی و سۆرانی ئەم ئەرکانە جێبەجێ دەکا:
  - پێشپرۆسەس کردنی دەق:
    "راستە لە وولاتی ٢٣هەمدا" ⇐ "ڕاستە لە وڵاتی 23هەمدا"
  - ڕاست کردنەوەی هەڵەی ڕێنووس:
    "سوتاندبووت" ⇐ ["ستاندبووت", "سووتاندبووت", "سووڕاندبووت"]
  - وەرگۆڕانی ئەلفوبێ:
    "rojhilata navîn" ⇐ "ڕۆژهلاتا ناڤین"
  - دۆزینەوەی ستێم:
    "دیتبامن" ⇐ "دی"
  - دۆزینەوەی لێما:
    "دیتبامن" ⇐ "دیتن"
  - ئانالیزی مۆرفۆلۆژی:
    "دیتبامن" ⇐ 'pos': ['verb'], 'base': 'دیت', 'suffixes': 'بامن'
  - لێکدانەوەی وشە:
    "ئەوان ڕێک کەوتن" ⇐ ["ئەوان", "ڕێک کەوتن"]

Kurdî NLP کوردی

تەکنۆلۆژیای زمان بۆ کوردی: چ بکرێ؟



سەرچاوەی وێنە: [9]

دەرنجامەکان

دەرنجامەکان

- پرۆژەکان بە شێوەیەکی کراوە بڵاو بکردرێنەوە
- بڕوامان بە کاری بەکۆمەڵ بێ نەک تەنیا تاقەکەسی
- هاندانی خوێندکاران و لێکۆڵێنەران بۆ ڕە چاو گرتنی ئەم بواری توێژینەوەیە

خولی زمانەوانیی کۆمپیوتەری: https://sinaahmadi.github.io/KurdishCL/

- بەرپرسایەتیی پسپۆڕانەمان لە حاندی زمانەکەماندا هەبێ
  - ئەگەر زمانەوانی، فێری زمانەوانیی کۆمپیوتەری بە
  - ئەگەر کۆمپیوتەرزانی، فێری پرۆسەس کردنی زمانی سروشتی بە
  - ئەگەر هیچیان نی، یارمەتی زمانەکە لە سەر تۆڕەکۆمەڵایەتییەکان بە
  - هەر کوردی نووسینیش یارمەتییەکی زۆرە بۆخۆی!

[1] How many languages are there in the world?
https://www.ethnologue.com/guides/how-many-languages.
Accessed: 2022-04-30.

[2] Unsupervised cross-lingual representation learning.
https://ruder.io/unsupervised-cross-lingual-learning.
Accessed: 2022-07-30.

[3] David Crystal.
Language death.
Cambridge University Press, 2002.

[4] Sina Ahmadi.
Building a corpus for the Zaza–Gorani language family.
In *Proceedings of the 7th Workshop on NLP for Similar Languages, Varieties and Dialects*, pages 70–78, 2020.

[5] Sina Ahmadi.
KLPT–Kurdish Language Processing Toolkit.
In *Proceedings of Second Workshop for NLP Open Source Software (NLP-OSS)*, pages 72–84, 2020.

[6] ST Baban and S Husein.
Programmable grammar of the kurdish language.
ILLC Research Report and Technical Notes, 1995.

[7] Géraldine Walther and Benoît Sagot.
Developing a large-scale lexicon for a less-resourced language: General methodology and preliminary experiments on sorani kurdish.
In *Proceedings of the 7th SaLTMiL Workshop on Creation and use of basic lexical resources for less-resourced languages (LREC 2010 Workshop)*, 2010.

[8] Sina Ahmadi.
A rule-based kurdish text transliteration system.
ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP), 18(2):1–8, 2019.

[9] Sina Ahmadi.
On the Current State of Kurdish Language Processing.
Proceedings of the 5th International Conference on Kurdish Linguistics (ICKL-5) Conference, 2021.

[10] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin.
Attention is all you need.
Advances in neural information processing systems, 30, 2017.

[11] Sepp Hochreiter and Jürgen Schmidhuber.
Long short-term memory.
Neural computation, 9(8):1735–1780, 1997.

[12] Alan M Turing.
Computing machinery and intelligence.
In *Parsing the turing test*, pages 23–65. Springer, 2009.

[13] John R Pierce and John B Carroll.
Language and machines: Computers in translation and linguistics, 1966.

[14] Noam Chomsky.
Syntactic structures.
In *Syntactic Structures*. De Gruyter Mouton, 2009.

[15] Karen Sparck Jones.
Natural language processing: a historical review.
Current issues in computational linguistics: in honour of Don Walker, pages 3–16, 1994.

[16] George Yule.
The study of language.
Cambridge university press, 2020.

[17] Derek Bickerton.
Adam's tongue: how humans made language, how language made humans.
Macmillan, 2009.

[18] Noam Chomsky.
Of minds and language.
Biolinguistics, 1:009–027, 2007.

[19] Ralph W Fasold and Jeff Connor-Linton.
An introduction to language and linguistics.
Cambridge university press, 2014.

Kurdî NLP کوردی