Information retrieval

لە ئەلگۆریتمەکەی ئێمەدا وەرگرتنی داتاو زانیاری لە ڕێگەی داونلۆد کردنی کۆمەڵێك فایلی ناو وێبسایتەکانەوە ئەبێت کە ئەم فایلانە

بە شێوەیەکی نیمچە ڕێکخراو ڕێكخراون وە داتاکانی ناوی سترەکچەرێکی ئاسانی نیە بۆ وەرگرتنی داتا لێیەوە بۆ نموونە داتای بە شێوەی خشتەیی یان لیست

ئاسانە وەرگرتنی داتا لێیانەوە، لەبەر ئەم هۆکارانە ئێمە هەڵساین بە داونلۆدکردنی تەواوی وێب پەیجەکە بۆ ئەوەی تەوەوی داتاکانی ناوی بەردەست بێت

بە پێچەوانەی هەندێک جۆری تری ئینفۆرمەیشن ڕیتریڤاڵەوە کە داتاکانیان بە شێوەی خشتەیی یان لیست دانراوەو بە ئاسانی داتا وەر ئەگرنەوە یان هەندێکیان

داتا لە داتابەیسەوە وەرئەگرنەوە کە کوێری بەکار ئەهێنن بۆ دۆزینەوەی داتای پێویست

لەم سەرچاوەی خوارەوەدا باسی ئینفۆرمەیشن ڕیتریڤاڵ بە گشتی کراوە وە باسی چەند ڕێگەیەک و ستراتیجی جیاواز کراوە بۆ وەرگرتنی داتا لە دۆکۆمێنتە ستاتیک و تەنانەت

ئەو دۆکۆمێنتانەی کە بە بەردەوامئ ئەگۆڕێن یان داتایان بۆ زیاد ئەبێت

Grossman, David A., and Ophir Frieder. *Information Retrieval: Algorithms and Heuristics*. Springer, 2004.

Information extraction

کاتێکی هەموو داتای ناو پەیجەکانمان بەو ئەلگۆریتمەی سەرەوە وەرگرت وە داتاکە بەردەستبوو بۆ ئەوەی ئەکسێسی بکەین ئەو داتایەی کە لامانە داتایەکی رێکنەخراوە وە هەندێک

شتی زیادە لە داتاکەیا هەیە کە پێویستمان پێی نیە وەک وێب تاگەکان بۆیە لەم ئەلگۆریتمەیان هەڵئەستین بە پارسکردنی داتاکە بۆئەوەی ئەو داتا زیادانەی کە پێویست نین لەیان بەرین

وە پاشان هەل ئەستین بە ڕێکخستنی داتاکە بۆ نموونە جیاکردنەوەی پۆستەکان و تێکستەکانی ناوی وە پاشان داتاکان هەموی بە شێوەی لیست یان ئەررەی ڕێک ئەخەین بۆ ئەوەی

لە هەنگاوی داهاتوا کە تێکست پرۆسێسینە بتوانین کاری لەسەر بکەین

بە هەمان شێوە لەم ڕیفرێنسەی خوارەوەیا بە شێوەیەکی ڕون باسی ئنفۆرمەیشن ئێکستراکشن کراوە بەشێوەیەکی گشتگیر وە باسی ڕێکخستنی داتای ڕێکنەخراو کراوە لەگەڵ

چەند تەکنیکو تیۆریەک وە هەروەها چەند کەیس ستەدیەک لەم کتێبەدا هەیە کە باس لە ڕێکخستنی داتای ناو نیوس و لێگاڵ تێکست و بایۆمێدیکاڵ داتا ئەکات

Moens, Marie-Francine. *Information Extraction: Algorithms and Prospects in a Retrieval Context*. Springer, 2006.

Natural language processing

لە دوای ئەوەی کە هەموو داتاکانمان ڕێکخست و داتا زیادەکانمان لەبرد ئێستا کاتی ئەوەیە کە ئەنەڵایزو پرۆسێسی ئەو داتایە بکەین کە بەدەستمانەوەیە بۆ بەدەستهێنانی ئامانجەکەمان

لە ئەلگۆریتمەکەی ئێمەدا دوو بەش لە داتامان هەیە کە پێویستا بەراوردیان بکەین بە یەک کە پێکهاتون لە ئەو وشانەی کە بەدوایان ئەگەڕێین کە لیستێکن لە وشەی تایبەت وە ئەو دۆکومێنتەی کە لا پەیجەکان

وەرمان گرتوە بۆیە یەتوانین بڵیین کە ئەلگۆریتمەکەمان وەکو سێرچیک کار ئەکات بۆ دۆزینەوەی کۆمەڵیک پاتێرنی تایبەت وە لەسەر ئەم پاتێرنانە بڕیار ئەیات و پەیجەکان جیا ئەکەتەوە کە چ جۆرێکن

وە لەم پەیپەری خوارەوەدا باسی ناچڕاڵ لانگویج پرۆسێسین کراوە بە هەمان شێوە ئەوەنیش داتاسێتیان بەکار ‌‌هێناوە بۆ بەڕێوەبردنی کاراکانیان کە داتا سێتەکان پێک هاتون لە داتای سپیچی چەند کەسێکو واتا

داتاکان تێکستن، لەگەڵ ئەوەشدا باسی چەند ڕێگاو تەکنیکێک کراوە بۆ دۆزینەوەی نەیم ئێنتیتی لەناو تێکستەکاندا کە نزیکە لە تەکنیکەکانی ئەلگۆریتمەکەی ئێمەوە

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural Language Processing (Almost) from Scratch. *NEC Laboratories America*