ئىنتوناتسىيە قائىدىلىرىنى ئاپتوماتىك ئايرىشقا قارتا ئېلىپ بېرىلغان ئۇيغۇر تىلى سۆز تۈركۈملىرىنى ئاپتوماتىك بەلگىلەش تېخنىكىسى تەتقىقاتى

نۇربىيە تاھىر[1]،دىلمۇرات تۇرسۇن[1*]،ئەسقەر روزى[2]

1(شىنجاڭ ئۇنىۋېرسىتېتى ئۇچۇر پەن-تېخنىكىسى ۋە ئىنژىنېرلىقى ئىنستىتوتى،ئۈرۈمچى 830046)

2(شىنجاڭ ئۇنىۋېرسىتېتى ماتېماتىكا ۋە سىستېما پەن تېخنىكىسى ئىنستىتوتى ،ئۈرۈمچى 830046)

ئۆزۈندە:بۇ ماقالدە تىل بېرىكتۈرۈش سىستېمىسىنىڭ تىكىست ئانالىز قىلىش مودىلىدىكى ئىنتوناتسىيە چېگرىسىنى ئاپتوماتىك ئايرىش تېخنىكا بۆلۈكى ئارقا كۆرۈنۈش قىلىنىپ،ئۇيغۇر تىلىدىكى سۆز تۈركۈملىرىنى ئاپتوماتىك بەلگىلەش تېخنىكىسى نوقتىلىق تەتقىق قىلىنىدى.يەنى:ئالدى بىلەن قوللىنىش ساھەسىدىكى ئالاھىدىلىكىگە ئاساسەن سۆز تۈركۈملىرىنىڭ سانى ۋە بېكتىش پىرىنسىپى بەلگىلىنىپ جۈملىلەر تاللاپ چىقىلدى ھەمدە ئۇنۇڭغا قارتا ئادەم كۈچى ئارقىلىق ھەر بىر سۆزنىڭ جۈملە مۇھىتىدىكى سۆز تۈركۈمى بەلگىلەندى؛ ئاندىن ستاتىستىكا ئۇسۇلى ئارقىلىق سۆز تۈركۈملىرىنىڭ نىسپى جەدۋىلى ۋە سۆز تۈركۈملىرىنىڭ ماسلىشىش جەدۋىلى ھاسىل قىلىندى؛ ئاخىردا HMM مودىلىنىڭ ئىككى ئىلمىنىتلىق ئۇسۇلى ئارقىلىق ئۇيغۇرتىلىنىڭ سۆز تۈركۈمىنى ئاپتوماتىك بەلگىلەش تېخنىكىسى ئەمەلگە ئاشۇرۇلدى.تەجرىبە جەريانىدا قوللىنىلغان ئالگورىتىمنىڭ ئۇنۇمىنى تەكشۈرۈش ئۈچۈن 10000 جۈملە مەشقىلەندۈرۈلۈپ،توپلام ئىچى ۋە سىرتىدىن ئايرىم-ئايرىم ھالدا 500 دانەدىن جۈملە تاللاپ ئېلىندى ھەمدە ئۇلارنىڭ سۆز تۈركۈملىرى ئاپتوماتىك بەلگىلىنىپ ئۇنىڭ توغرۇلۇق دەرىجىسى تەكشۈرۈلدى.تەجرىبە نەتىجىسى ئانالىز ئۇسۇلىنىڭ قوللىنىش قىممىتى بارلىقىنى دەللىلەپ بەردى.

ھالقىلىق سۆزلەر: ئۇيغۇرتىلى؛ ئىنتوناتسىيە قائىلىمى؛ سۆز تۈركۈمىنى بەلگىلەش؛ تىل بېرىكتۈرۈش؛ تىكىست ئانالىزى

Research on a POS tagging method for detecting the boundary of prosody in Uyghur Sentences

Nurbiye Tahir[1]， Dilmurat Tursun[1*]， A skar Rozi[2]

[1]（College of Information Science and Engineering, Xinjiang University, Urumqi 830046, China）

[2]（College of Mathematics and System Science, Xinjiang University, Urumqi 830046, China）

[Abstract]: As a critical issue in the text analysis part of Uyghur synthesis system, an automatic POS (part of speech) tagging method for estimation the boundary of different prosody levels in text sentences is mainly discussed in this paper. At first, according to the characteristics of specific application field, the category of parts used in POS tagging and the rules are confirmed, candidate sentences are selected and are tagged by manually, then POS probability list tables and relative occurring position information tables of POS are achieved by statistics, and at last, the automatic POS tagging method is implemented through adopting bigram model on the basis of HMM model. In order to approve the validity of the method presented in this paper, the large scale text corpus is used in which 10000 sentences are selected for training and 500 sentences are used for testing. Test results show that the method adopted in this paper for Uyghur POS tagging is feasible, appropriate and valid.

[Keywords]: Uyghur; prosodic layer; POS tagging; speech synthesis; text analysis

1. كىرىش سۆز

2. سۆز تۈركۈمى نىسپەت جەدۋىلى

يېقىنقى يىللاردىن بۇيان،ئاز سانلىق مىللەت تىل-يېزىق ئۇچۇر بىر تەرەپ قىلىش ساھەسىدىكى يادرولۇق تېخنىكىلارنىڭ بىرى بولغان_ئادەم ۋە ماشىنا ئالاقىسى تېخنىكىسى،ئۇيغۇر تىل بېرىكتۈرۈش تەتقىقاتىدا ئاساس بولۇش نەزىرىيە تەتقىقاتى ۋە تېخنىكا قوللۇنۇلۇش قاتارلىق تەرەپلەردە كۆرنەرلىك ئىلگىرىلەشلەرگە ئېرىشتى.ئەمما بېرىكتۈرۈلگەندىن كېيىنكى ئاۋازنىڭ تەبئىيلىكى يەنىلا كۆزلىگەن نىشاندىن خېلىلا يىراق بولماقتا.بېرىكتۈرۈلگەن ئاۋازنىڭ تەبئىيلىكىنى تېخمۇ يۇقىرى كۆتۈرۈش ئۈچۈن ئاۋاز ئامبىرىنى قايتىدىن قۇرۇش،مەسىلەن توپلىمىنى ئەلالاشتۇرۇش،تىل بېرىكتۈرۈش جەريانىدىكى ئالدىنقى بۆلەك مودىل قۇرلىشىنى ئىنچىكە لايىھەلەش،ئىنتوناتسىيەلىك بۆلەكلەر چېگرىسىنى ئالدىن پەرەز قىلىپ بەلگىلەش، ئۇرغۇنى ئالدىن پەرەز قىلىش،سۆز

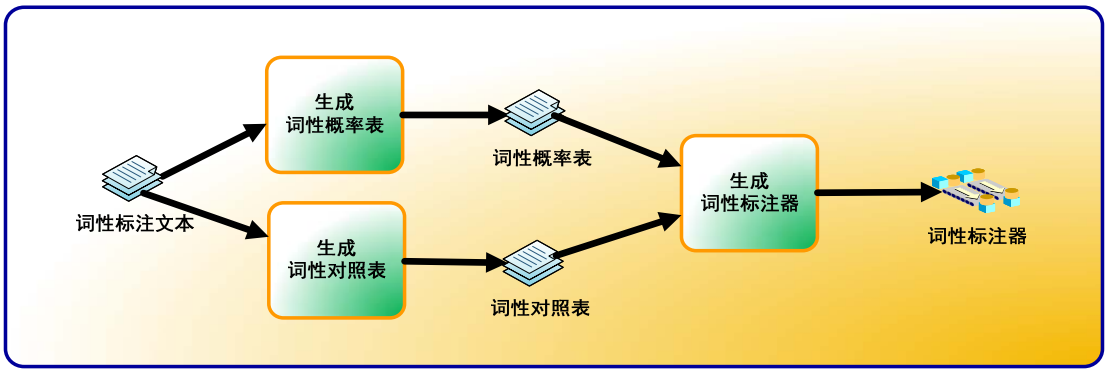
تۈركۈمىنى ئالدىن بەلگىلەش ئۇسۇلىنى ئەلالاشتۇرۇش قاتارلىق تەرەپلەردە يەنمۇ بىر قەدەم ئىلگىرلىگەن ئاساستا تەتقىق قىلىشقا ۋە ئىزدىنىشكە توغرا كېلىدۇ.شۇڭلاشقا ئېنتوناتسىيە بۆلەكلىرىنىڭ چېگرىسىنى ئالدىن پەرەز قىلىپ بەلگىلەش بۇ ساھەدىكى قىزغىن نوقتىغا،شۇنداقلا بىرىكتۈرۈلگەن ئاۋازنىڭ تەبئىيلىك دەرىجىسىنى يۇقىرى كۆتۈرۈشتىكى مۇھىم ھالقىلىق مەسىلىگە ئايلاندى .

ئادەتتە،تىل بىرىكتۈرۈش سىستېمىسى تۆۋەندىكىدەك تۆت قىسىمنى ئۆز ئىچىگە ئالىدۇ:تېكىست ئانالىز قىلىش مودىلى،ئېنتوناتسىيە ھاسىل قىلىش مودىلى،ئاۋاز ئىلمى مودىلى،ئاۋاز ئاساسى بىرلىكى ۋە ئاۋاز پارامېتىرى ئامبىرى مودىلىدىن ئىبارەت[1].ئېنتوناتسىيە قەۋەتلىرىنى ئايرىش جەريانىدا،كىرگۈزۈلگەن تېكىستنىڭ خاس ئۇچۇرلىرى بولۇش بىلەن بىرگە تېكىستكى ھەر قايسى سۆزلەرنىڭ سۆز تۈركۈمى " ئاپتۇماتىك سۆز تۈركۈمى بەلگۈلىگۈچ" ئارقىلىق ئېرىشىلگەن بولىشى كېرەك[2]. ئۇيغۇرچە تىل بىرىكتۈرش سىستېمىسىدىكى ئېنتوناتسىيەلىك تەركىبلەرنىڭ چېگرىسىنى ئاپتۇماتىك ئايرىش بولسا تېكىست ئانالىز قىلىش مودىلىدىكى مۇھىم بىر ھالقا بولۇپ ھېسابلىنىدۇ. ئېنتوناتسىيەلىك تەركىبلەرنىڭ چېگرىسىنى ئايرىشنىڭ توغرۇلۇق دەرىجىسىنى ئاشۇرۇش ئۈچۈن ئەڭ ئۇنۇملۇك بولغان سۆز تۈركۈمىنى ئاپتۇماتىك بەلگىلەش ئۇسۇلىنى تاللاپ قوللىنىش كېرەك.بۇ خىزمەتنىڭ ۋەزىپىسى بولسا،كونكرېتنى تىل مۇھىتىدە ھەر بىر سۆزگە ئېنىق بىر سۆز تۈركۈمىنى بەلگىلەپ چىقىشتىن ئىبارەت. بۇ جەرىم دەرىخى مودىلى بىلەن ماسلىشىپ ئېنتوناتسىيەلىك تەركىبلەرنىڭ چېگرىسىنى ئاپتۇماتىك ئايرىش جەريانىدا مۇھىم رول ئويناىدۇ. ئۇيغۇرتىلىدىكى ئېنتوناتسىيەلىك تەركىبلەرنىڭ چېگرىسىنى ئاپتۇماتىك ھالدا توغرا ئايرىپ چىقىش بولسا،ئەڭ ئاخىردا تىل بىرىكتۈرش سىستېمىسىدا بىرىكتۈرۈلگەن ئاۋازنىڭ تەبئىيلىكىنى يۇقىرى كۆتۈرۈشكە خىزمەت قىلىدۇ. شۇڭلاشقا ئېرىشمەكچى بولغان نەتىجە ئۇيغۇرچە تىل بىرىكتۈرۈش تەتقىقاتىدا ناھايىتى مۇھىم ئورۇندا تۇرىدۇ.

بۇ خىزمەت جەريانىدا بەلگىلەنگەن سۆز تۈركۈمىنىڭ توغرىلىقىغا كاپالەتلىك قىلىش ئۈچۈن ئالدى بىلەن 10610 جۈملىگە قارتا خاس مۇھىتتىكى سۆز تۈركۈمىنى بەلگىلەش خىزمىتى ئېلىپ بېرىلدى.بۇ سانلىق مەلۇماتلار ئۇيغۇرتىلى ئېنتوناتسىيە تەتقىقات خىزمىتىنىڭ چۇڭقۇرلاپ قانات يايدۇرلىشى ئۈچۈن ناھايىتى ياخشى ئاساس سېلىپ بەردى؛بۇ يەنە ئۆز نۆۋىتىدە،بىرىكتۈرۈلگەندىن كېيىنكى ئاۋازنىڭ تەبئىيلىك دەرىجىسىنى ئاشۇرۇش،پۈتكۈل سىستېمىنىڭ تېكىست ئانالىز قىلىش مودىلىدىكى تېخنىكىلىق خىزمەتلەرنىڭ چۇڭقۇرلاپ قانات يايدۇرلىشى ئۈچۈنمۇ ئاساس بولغۇسى! بۇ خىزمەت جەريانىدا 10610 جۈملىدىن 10000 جۈملە تاللىۋېلىنىپ مەشقىلەندۇرۇلدى،توپلام ئىچى ۋە سىرتىدىن ئايرىم-ئايرىم ھالدا 500 جۈملە تاللىۋېلىنىپ سىناق قىلىنىدى.بۇ تۈردىكى تەتقىقاتلار ئاۋاز بىر تەرەپ قىلىش تەتقىقاتىنىڭ قوللىنىش ساھەلىرىدە،يەنى:تىل بىرىكتۈرۈش،تىلنى تونۇش،ئاۋاز تەتقىقاتى،تىل تەتقىقاتى قاتارلىق ساھەلەردە ئىلمى ۋە ئەمەلىي قىممىتى بار.بۇنىڭ يەنە ئۇيغۇرتىلى ھەتتا پۈتكۈل ئالتاي تىللىرى سىستېمىسىدىكى تىللارنىڭ ئاۋاز تەتقىقاتى قوللىنىش ۋە ئېچىش جەريانىدىمۇ خېلى يۇقىرى پايدىلىنىش قىممىتى بار.

2. سۆز تۈركۈمىنى ئاپتۇماتىك بەلگىلەش سىستېمىسىنىڭ مودىللاشقان لايىھەسى

بۇ خىزمەتتىكى سۆز تۈركۈمىنى ئاپتۇماتىك بەلگىلەش سىستېمىسى:سۆز تۈركۈمىنى بەلگىلەش تېكىست مودىلى،سۆز تۈركۈمى ستاتىستىكا ئۇچۇرلىرىغا ئېرىشىش مودىلى،سۆز تۈركۈمىنى ئاپتۇماتىك بەلگىلىگۈچ مودىلىدىن تەركىپ تاپقان.1-رەسىمدە كۆرسىتىلگەندەك:



1-رەسىم سۆز تۈركۈمىنى ئاپتۇماتىك بەلگىلەش سىستېمىسىنىڭ قۇرۇلمىسى

2.1 سۆز تۈركۈمىنى بەلگىلەش تېكىست مودىلى

2.1.1 سۆز تۈركۈمى ۋە ئۇنىڭ سانى

سۆز تۈركۈمىنى بەلگىلەش ئېنگلىزچىدا (Part-of-Speech tagging OR POS tagging) دېيىلىدۇ.بۇ ئامباردىكى ھەر بىر سۆزگە ماس بولغان تىپ يەنى سۆزلەرنىڭ تەۋەلىك گۇرۇپپىسىنى بەلگىلەپ چىقىش جەريانىدىن ئىبارەت. گەرچە تەبئىي تىل-ئالاقە جەريانىدىكى بەلگىلەر تېخمۇ كۆپ مەنىلەرگە ئىگە بولۇشتەك ئالاھىدىلىككە ئىگە بولسىمۇ،ئەمما تەبئىي تىلنىڭ ئالاقە جەريانىدىكى ئالاھىدىلىكلىرىنى بەلگىلەش جەريانى ماشىنا تىلىدىكى سۆزلەرنىڭ ئەسلىگە قايتۇرلۇش (tokenization) جەريانى بىلەن ئوخشىشىپ كېتىدۇ[2].

سۆز تۈركۈمىنىڭ تىل ئۇچۇرلىرىنى بىر تەرەپ قىلىشتىكى ئەھمىيىتى-سۆز ۋە ئۇنىڭ ئەتراپىدىكى بۆلەكلەرگە مۇناسىۋەتلىك كۆپلىگەن ئۇچۇرلارنى تەمىنلەپ بېرەلىگەنلىكىدە.شۇڭلاشقا سۆز تۈركۈمىنى ئاپتۇماتىك بەلگىلەش ئاۋاز پەرقلەندۈرۈش،تەبئىي تىلنى تەھلىل قىلىش ۋە ئۇچۇر ئىزدەش قاتارلىق ساھەلەردە بارغانسىرى كۆپ قوللىنىلماقتا.سۆز تۈركۈملىرى بەلگىلىنىپ بولغان ماتېريال ئامبىرى تىل تەتقىقاتىدا ناھايىتى مۇھىم ئەھمىيەتكە ئىگە.تىل بېرىكتۈرۈش سېستېمىسىدا ئەگەر بىر سۆزنىڭ سۆز تۈركۈمىنى بىلسەك تېخمۇ تەبئىي بولغان ئاۋازغا ئېرشەلەيمىز.ئۇيغۇر تىلىدىكى سۆز تۈركۈملىرىنىڭ تۈرلىرى بولسا سۆز يىلتىزلىرىنىڭ گىرامماتىكىلىق ئالاھىدىلىكلەر بويىچە بۆلۈنۈشىدىن ئىبارەت.گىرامماتىكىلىق ئالاھىدىلىكلەر بولسا:شەكىل ئۆزگۈرۈش،سۆز بىلەن سۆزنىڭ بېرىكىش ئىقتىدارى،سۆزنىڭ جۈملىدىكى ئىقتىدارى ۋە سۆزنىڭ قۇرۇلۇش ئالاھىدىلىكى قاتارلىقلاردىن ئىبارەت.ئوخشاش بىر تۈركۈمگە تەۋە بولغان سۆزلەر ئوخشاش گىرامماتىكىلىق ئالاھىدىلىككە ئىگە بولىدۇ.

ئىلگىرىكى ئۇيغۇرچە تىل بېرىكتۈرۈش سېستېمىسىنىڭ سۆز تۈركۈمىنى بەلگىلەش جەريانىدا 16 تۈرلۈك سۆز تىپى قوللىنىلغان[3]،بۇ خىزمەت جەريانىدا بېيىجىڭ ئۇنۋېرسىتېتىنىڭ سۆز تۈركۈمىنى بەلگىلەش قائىدىسى ۋە ئىلگىرى قوللانغان 16 خىل سۆز تۈركۈمى ئاساسىدا، ئېنتوناتسىيەگە كۆرسىتىدىغان تەسىرى بىر قەدەر چوڭ دەپ قارالغان 41 خىل سۆز تۈركۈمى دەسلەپكى قەدەمدە بېكىتىلدى.ستاتىستىكا خىزمىتى ئاخىرلاشقاندا سۆز تۈركۈمىنىڭ سانىنى مۇقۇملاشتۇرۇشقا قارتا يەنە بىر قېتىم جەزملەشتۈرۈش خىزمىتى ئىشلىنىپ ئەڭ ئاخىردا 37 خىل سۆز تۈركۈمى مۇقۇملاشتۇرۇلدى.بۇ سۆز تۈركۈملىرىنىڭ ئىپادىلىنىشى تۆۋەندىكى 1-جەدۋلدە قىسمى كۆرسىتىلگەندەك:

1-جەدۋل سۆز تۈركۈملىرى جەدۋلى

文标识	英文标识	维文标识	序号	中文标识	英语标识	维语标识	序号
不定代词	rb	بەلگىسىز ئالماش	20	名词	n	ئىسىم	1
反身代词	ro	ئۆزلۈك ئالماش	21	人名	nr	كىشى ئىسىملىرى	2
副词	d	رەۋىش	22	地名	ns	يەر-جاي نامى	3
情态副词	dh	ھالەت رەۋىش	23	位置名词	nt	ئورگان،تەشكىلات نامى	4
耐间副词	dw	ۋاقىت رەۋىش	24	行业术语	nz	كەسىپ ئاتالغۇلار	5
位置副词	do	ئورۇن رەۋىش	25	形容词	a	سۈپەت	6

2.1.2 تېكىست لايىھەسى

تېكىست لايىھەسىنىڭ مەقسىتى: ئامالنىڭ بارچە سان جەھەتتىن ئاز بولغان تېكىست ئارقىلىق تېخمۇ كۆپ سۆز تۈركۈمى ھادىسىنى ئۆز ئىچىگە ئېلىش.بۇنداق بولغاندا ھەر خىل جۈملە شەكىللىرى ۋە ھەر خىل سۆز تۈركۈملىرىنىڭ كۆرۈلۈش نىسپىتىگە كاپالەتلىك قىلغىلى بولىدۇ.

جۈملە تاللىغاندا جۈملە ئۇزۇنلىقى نەزەرگە ئېلىنىپ،جۈملە شەكلى بويىچە تۈرگە ئايرىپ تاللاندى.ئالدى بىلەن ئەسلىدىكى تېكىستنى بوغۇم،سۆز،سۆز بېرىكمىسى ۋە جۈملىلەرگە پارچىلاپ ئاندىن Greedy ئالگورىتمى ئارقىلىق زور كۆلەمدىكى ماتېريال ئامبىرىدىن ۋەكىللىك خاراكتېرىگە ئىگە بولغان جۈملىلەر توپلىمى ھاسىل قىلىندى[3].تېكىستنى تاللاپ بولغاندىن كېيىن ئۇنۇڭغا قارتا ئادەم كۈچى ئارقىلىق ئۆزگەرتىش،تاللاش ۋە شاللاش

خىزمىتى تەكرار ئېلىپ بېرىلىپ كەمتۈك جايلىرى تولۇقلاندى.

2.1.3 ئادەم كۆچى بىلەن سۆز تۈركۈمىنى بەلگىلەش

ئۇيغۇرتىلى يېپىشقاقلىقتا ئىگە بىر تىل بولۇپ،شەكىل ئۆزگۈرشى ناھايىتى مۇرەككەپ.ئۇنىڭ ئۈستىگە بىر قانچە سۆز تۈركۈمىگە تەۋە بولىدىغان سۆزلەر بىر قەدەر كۆپ،قوللىنىلىش نىسپىتىمۇ بىر قەدەر يۇقىرى[4].بۇ ماقالىدىكى سۆز تۈركۈمىنى بەلگىلەش ئاساسى-شۇ سۆزنى جۈملە مۇھىتىدە تەھلىل قىلىپ ئۇنىڭ سۆز تۈركۈمىنى بەلگىلەش.بۇ سەۋەپلىك ئادەم كۆچى ئارقىلىق بەلگىلىگەندە بەزىدە جۈملىدىكى شۇ سۆزنىڭ ئۆزىنىلا ئويلىشىشقا توغرا كەلسە بەزىدە سۆز تۈمۈرى،سۆز تۈرمۇرىنىڭ ئۆزگۈرشى،سۆز تۈمۈرى ۋە قوشۇمچىلارنىڭ ماسلىشىشى،ھەر خىل سۆز تۈرلىگۈچى قوشۇمچىلار ۋە سۆز ياسىغۇچى قوشۇمچىلارنىڭ ئۆز ئارا ماسلىشىشى قاتارلىق ئامىللارنى ئويلىشىشقا توغرا كېلىدۇ.شۇڭلاشقا بۇنۇڭغا قارتا بىرلىككە كەلگەن ئۆلچەم بېكىتىش قىيىن بولىدۇ.بەلگىلەنگەن سۆز تۈركۈمىنىڭ توغرىلىقىغا كاپالەتلىك قىلىش ئۈچۈن،بۇ گورۇپپىدىكى ئەزالار سۆزلەرنىڭ تۈركۈملەرگە بۆلۈنىشى ۋە مۇناسىۋەتلىك بىلىملەرگە پىششىق بولغانلار ۋە ئۆزۈن يىللىق ئۇيغۇرتىلى گىرامماتىكا ئوقۇتۇش تەتقىقاتى بىلەن شۇغۇللانغان ئۇقۇتقۇچىلاردىن تەشكىللەندى.

ئادەم كۆچى ئارقىلىق تەييارلانغان بۇ خىل ھۆججەتنىڭ مەزمۇنى تۆۋەندە كۆرسىتىلگەندەك:

1. ئۇيان/do بۇيان/doئۆتۈشۈۋاتقان/avشۇنچە/rkكۆپ/aئادەملەر/nئارىسىدا/doشەيخ/nيەنىلا/cئۇنىڭ/raئارقىسىدىن/nجىمغىنا/aقاراپ/dvتۇراتتى/v.
2. قىزمۇ/nكۈلمەك/nvبولۇۋىدى/vشەيخ/nسول/aكۆزىنى/nقىسىپ/dvئوڭ/aقېشىنى/nئاتتى/v.
3. ئۇ/raئەمدى/yئاتىسىغا/nئەگىشىپ/dvسالام/nقىلىشتىنمۇ/nvقالدى/v.
4. بۇ/rkئالا/aقانات/nقاپ/nيۇرەك/nيېگىتنىڭ/nقانداق/rsئىنتىقام/nئىلىش/nvئۈستىدە/nئويلاندى/v.

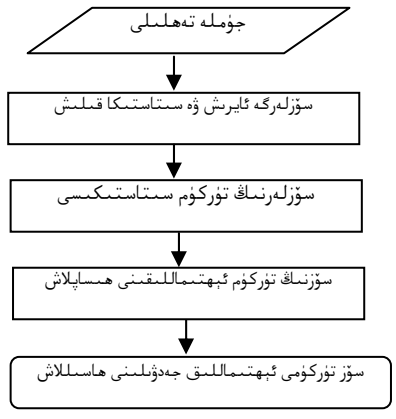
2.2 سۆز تۈركۈمى ستاتىستىكا ئۇچۇرلىرىغا ئېرىشىش مودىلى

سۆز تۈركۈمى ئېھتىماللىق جەدۋىلى ۋە سۆز تۈركۈملىرىنىڭ نىسپەت جەدۋىلى كىينىكى باسقۇچلاردىكى سۆز تۈركۈمىنى بەلگىلىگۈچ مودىلىدا ئىشلىتىلىدۇ.بۇ بۆلەكتە ئۇلارنىڭ ھىساپلاش ئۇسۇلى قىسقىچە تونۇشتۇرۇلىدۇ.

2.2.1 سۆز تۈركۈمى ئېھتىماللىق جەدۋىلى

VisualStudio2008 مۇھىتىدا C# پىروگىراممىا تىلى ئارقىلىق سۆز تۈركۈمى ئېھتىماللىق جەدۋىلىگە ئېرىشىمىز،بۇ جەدۋەلنى ھاسىل قىلىش جەريانىدا يېزىلغان يادرولۇق كود بۆلىكى تۆۋەندىكىچە:

```
foreach (Sentence sen in Sentences)
{foreach (Word wrd in sen.Words)
{if (TotalWords.ContainsKey(wrd.Content))//سۆزنىڭ كۆرۈلۈش قېتىم سانىنى ھىساپلاش
{TotalWords[wrd.Content]++;}
else
{TotalWords.Add(wrd.Content, 1);}
if (pools.Keys.Contains(wrd.Content +wrd.Turkum))//
ئەگەر نۆۋەتتىكى سۆز كۆرۈلگەن بولسا پەقەت ئۇنىڭ سۆز تۈركۈمىگە قارتا ھىساپلاش ئېلىپ بېرىش
{pools[wrd.Content + wrd.Turkum].Count++;}
else
{pools.Add(wrd.Content + wrd.Turkum, new WordPool(wrd, 1));}}//
يېڭىدىن كۆرۈلگەن سۆز ۋە ئۇنىڭ سۆز تۈركۈمىنى خاتىرلەش
nCount++;
```



2-رەسىم سۆز تۈركۈمى ئېھتىماللىق جەدۋىلى دىئاگىراممىسى

ھاسىل قىلىنغان سۆز تۈركۈمى ئېھتىماللىق جەدۋىلىنىڭ مەزمۇنى ۋە ھىساپلاش ئۇسۇلى 2-رەسىمدە كۆرسىتىلگەندەك:

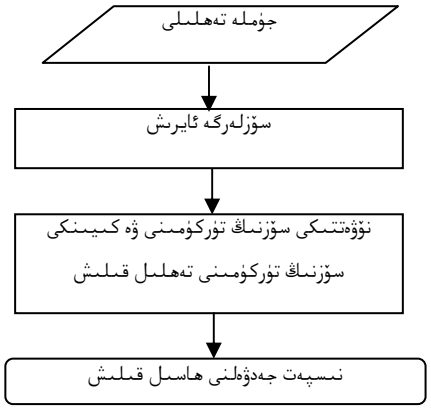
```
ProbWord[]={{"ikki", "m", 0.0020703933747422},
{"dOrE", "q", 0.5},
{"mExrEptin", "n", 0.5},
{"keyin", "dv", 0.00429184549356223},
```

{"HEr","re",0.00806451612903226},
{"ikki","m",0.0020703933747422},
……

2.2.1 سۆز تۈركۈمى نسپەت جەدۋلى

VisualStudio2008 مۇھىتىدا تۈرۈپ C# تىلى ئارقىلىق پروگراممىا تۈزۈش ئارقىلىق سۆز تۈركۈملىرى بەلگىلەنگەن تېكىستتىن سۆز تۈركۈمى نسپەت جەدۋلىگە ئېرىشىمىز،بۇ جەدۋەلنى ھاسىل قىلىش جەريانىدا يېزىلغان يادرولۇق كود بۆلىكى تۆۋەندىكىچە:

```
for (i = 0; i < 38; i++)
{sb.Append("{");
sbHtm.AppendLine("<tr>");
for (j = 0; j < 38; j++)
{n4[i, 38] = n4[i, 38] + n4[i, j]; }//هەر قايسى سۆز تۈركۈملىرىنىڭ كۆرلۈش قېتىم سانىنى ھىساپلاش
nTotal = nTotal + n4[i, 38];
for (j = 0; j < 39; j++)
{sb.Append(n4[i, j]);
sbHtm.AppendLine("<td>" + n4[i, j] + "</td>");
if (j < 38)
{sb.Append(",");}}
```



3-رەسىم سۆز تۈركۈملەرنىڭ نسپەت جەدۋلى دىئاگراممىسى

ھاسىل قىلىنغان سۆز تۈركۈمى نسپى جەدۋلىنىڭ مەزمۇنى ۋە ھىساپلاش ئۈسۈلى 3-رەسىمدە كۆرسىتىلگەندەك:

```
Int BiGramProbs[37][38]={{0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0}
{0, 987, 16, 173, 39, 56, 75, 43, 29, 4, 161, 8, 1, 5, 8, 5016, 92,
123, 3, 56, 0, 1, 53, 5, 105, 10, 68, 24, 5, 8, 18, 32, 1, 881, 578,
175, 427, 44, 32, 0, 1, 0, 9363}
……
```

2.3 سۆز تۈركۈمىنى ئاپتۇماتىك بەلگىلىگۈچى مودىلى

ئاپتۇماتىك بەلگىلەشنى ئەمەلگە ئاشۇردىغان ئالگورىتىملارنى ئىككى تۈرگە ئايرىشقا بولىدۇ،ئۇلار:مەلۇم قائىدىلەر ئاساسىدىكى ئۈسۈل （rule-based tagger） ۋە رەندىم ئايرىش ئۈسۈلى （stochastic tagger）.رەندىم ئايرىش ئۈسۈلىدا ئادەتتە مەشقىلەندۈرۈش ئامبىرى ئارقىلىق،بەلگىلەش ئوبىكتى بولغان ھەر بىر سۆزنىڭ جۈملە مۇھىتىدىكى تەۋەلىك تىپ ئېھتىماللىقى ھىساپلاپ چىقىلىدۇ.مەسىلەن: HMM بەلگىلەش سىستېمىسى[2].

بۇ ماقالىدە ئىنتوناتسىيە قەۋەتلىرىنى ئاپتۇماتىك ئايرىش ئېھتىياجىغا ئاساسەن HMM بەلگىلەش مۇدىلىدىكى ئىككى ئىلمىنتىلىق ئۈسۈل قوللۇنۇلۇپ ئۇيغۇرتىلى سۆز تۈركۈمىنى ئاپتۇماتىك بەلگىلەش سىستېمىسىنىڭ ئاساسى گەۋدىسى ھاسىللاندى. HMM بەلگىلەش مۇدىلىدىكى ئىككى ئىلمىنتىلىق ئۈسۈلنىڭ ھىساپلاش جەريانى تۆۋەندىكىچە:

1.) "." بەلگىسى بولسا جۈملە ئاخىرلاشقانلىقىنىڭ ئىپادىسى بولۇپ،بۇنىڭدىن پايدىلىنىپ بىر جۈملە تاللاپ ئېلىۋېلىندۇ.

2.) مەشقىلەندۈرۈلىدىغان جۈملىلەردىن ھەر قايسى جۈملىلەر ئايرىم-ئايرىم تاللاپ ئېلىنىپ بۇ جۈملىلەرنى تەشكىللىگۈچى سۆزلەرنىڭ سۆز تۈركۈمى نسپىتى ھىساپلاپ چىقىلىدۇ.

$$P(w_i \mid c_i) = \frac{\text{训练语料中}\quad w_i\text{的词性被标记为}\quad c_i\text{的次数}}{\text{训练语料中}\quad c_i\text{出现的总次数}} \qquad (1)$$

فورمۇلاسى :

سۈرەت: مەشقىلەندۈرۈش تۇپىدا w_i سۆزنىڭ سۆز تۈركۈمى c_i دەپ بەلگىلىنىش قېتىم سانى

مەخرەج: مەشقىلەندۈرۈش تۇپىدا c_i نىڭ جەمئىي كۆرۈلۈش قېتىم سانى

3.) نۆۋەتتىكى سۆزنىڭ ئالدىدا كۆرۈلگەن بارلىق سۆز تۈركۈملىرى تارتىپ چىقىرتىلىپ تۇرغان نوقتا ھاسىل قىلىنىدۇ،ئاندىن ھەر بىر سۆز تۈركۈمىنىڭ ئۆزىنىڭ ئالدىدىكى سۆزنىڭ ھەر بىر سۆز تۈركۈمى بىلەن بولغان بارلىق گۇرۇپپىلىنىش ئېھتىماللىقى ھىساپلىنىدۇ،ھەمدە بۇ ئارقىلىق بىز يۇقاردا تونۇشتۇرۇپ ئۆتكەن سۆز تۈركۈملىرىنىڭ نىسپى جەدۋىلى ھاسىللىنىدۇ.

$$P(c_i \mid c_{i-1}) = \frac{\text{训练语料中 } c_i\text{出现在 } c_{i-1}\text{之后的次数}}{\text{训练语料中 } c_{i-1}\text{出现的总次数}} \qquad (2)$$

فورمۇلاسى :

سۈرەت: مەشقىلەندۈرۈش تۇپىدا c_i نىڭ c_{i-1} دىن كېيىن كېلىدىغان ئەھۋالنىڭ كۆرلۈش قېتىم سانى

مەخرەج: مەشقىلەندۈرۈش تۇپىدا c_{i-1} نىڭ جەمئىي كۆرۈلۈش قېتىم سانى

4.) سۆزنىڭ تۈركۈمىنى بېكىتىشتە،جۈملىدىكى سۆزلەرنىڭ تەرتىپى بويىچە باشلىنىش نوقتىسىغا باشلىنىش تۈگۈن نوقتىسىنى قويۇۋالىمىز.بىرىنچى سۆزنىڭ بارلىق ئېھتىماللىق سۆز تۈركۈملىرىنى تارتىپ چىقىرۇۋېلىپ ھەر بىر سۆز تۈركۈمىگە قارتا تۈگۈن نوقتىسى قۇرۇۋالىمىز،بىرىنچى سۆزنىڭ ھەر بىر سۆز تۈركۈمى تۈگۈن نوقتىسىنى باشلىنىش تۈگۈن نوقتىسىغا يۈزلەندۈرۈپ كىينكى سۆزنى چىقىرۇۋالىمىز.

5.) HMM ئاساسىدىكى بەلگىلەش ئالگورىتمىدا سۆزنىڭ ئىككى ئىلمىنتلىق ئېھتىماللىقى بىلەن گورۇپپىلىنىش ئېھتىماللىقىنىڭ كۆپەيتمىسى ئەڭ چوڭ بولغان بەلگىلەر ئارقمۇ-ئارقىلىقى ئەڭ ئاخىرقى يەكۈن قىلىنىدۇ[2].يەنى $P(w_i \mid c_i)P(c_i \mid c_{i-1})$ ھىساپلىنىپ،نەتىجىسى ئەڭ چوڭ بولغان ئىككى ئىلمىنتلىق سۆز تۈركۈمى گورۇپپىسى تېپىپ چىقىلىدۇ،ھەمدە نۆۋەتتىكى سۆزنىڭ سۆز تۈركۈمى تۈگۈن نوقتىسىنىڭ ئالدىدىكى ئەڭ ئۇنىملىك تۈگۈن نوقتىسى بىلەن بىرلىكتە،بۇ ئىككى ئىلمىنتلىق گورۇپپىنىڭ كەينىدىكى سۆز تۈركۈمى تۈگۈن نوقتىللىرىنى كۆرسىتىدۇ.

6.) ئەگەر بىر تەرەپ قىلىنمىغان سۆز بولسا بەشىنچى باسقۇچ بويىچە داۋاملىق ھىساپلاش ۋە بەلگىلەش ئېلىپ بېرىلىدۇ،ئەگەر ئۇنداق سۆزلەر بولمىسا كىينكى باسقۇچ ئورۇندىلىدۇ.

7.) بارلىق سۆزلەر بىر تەرەپ قىلىنىپ بولۇنغاندىن كېيىن جۈملىنىڭ ئەڭ كەينىدىكى بىر سۆزدىن باشلاپ ھەر بىر سۆزنىڭ ئۆزىگە خاس بولغان بارلىق سۆز تۈركۈملىرى ئىچىدىكى ئەڭ مۇۋاپىق بولغان سۆز تۈركۈمى لىنىيەسى بويىچە بۇ جۈملىدىكى ھەر بىر سۆزنىڭ سۆز تۈركۈمى ئارقمۇ-ئارقىلىقى تېپىپ چىقىلىدۇ.

VisualStudio2008 مۇھىتىدا تۇرۇپ C# تىلى ئارقىلىق پىروگىراممىا تۈزۈش ئارقىلىق بۇ بۆلەك ئاچقۇچلۇق ئالگورىتىم بۆلۈكىدىن ئازراق قىلغىنىمىزدىكىدەك نەتىجىگە ئېرىشىمىز،بۇ جەريانىدا يېزىلغان يادرولۇق كود بۆلىكى تۆۋەندىكىچە:

```
for (int nIndex = straWords.Count() - 1; nIndex > -1; nIndex--)
{if (nIndex == straWords.Count() - 1)
{if (lstWords.Where(w => w.Content == straWords[nIndex]).Count() > 0)
{var WordResult = lstWords.Where(w => w.Content == straWords[nIndex]).GroupBy(w =>
w.Turkum).OrderByDescending(w => w.Count()).First();
```

```
straResultWords[nIndex] = straWords[nIndex] + "/" + WordResult.Key;}
else
{straResultWords[nIndex] = straWords[nIndex] + "/n";}}
else
{List<NisbetJedwili> lstCurrentCollection = new List<NisbetJedwili>();
var CurrentColection = lstNisbet.Where(n => n.Content == straWords[nIndex]);
foreach (var c in CurrentColection)
{lstCurrentCollection.Add(c);}
List<NisbetJedwili> lstNextCollection = new List<NisbetJedwili>();
var NextColection = lstNisbet.Where(n => n.Content == straWords[nIndex + 1]);
foreach (var n in NextColection)
{lstNextCollection.Add(n);}
straResultWords[nIndex]   =   straWords[nIndex]   +   getTurkum(lstCurrentCollection,
lstNextCollection);}}
```

3.تەجرىبە مۇھىتى ۋە ئانالىزى

3.1 سانلىق مەلۇماتلارنى تەييارلاش

تەجرىبە سانلىق مەلۇماتلىرىنى ھازىرلاش جەريانىدا مەشىقلەندۈرۈش سانلىق مەلۇماتلىرى ۋە سىناق سانلىق مەلۇماتلىرى ھازىرلاندى.بۇنۇڭدا گېزىت-ژورنال،ئەدەبىي ئەسەرلەر قاتارلىق مەدىئا ۋاستىللىرىدىن تىكىست ماتېريالى يىغۇېلىنىپ ئاندىن تىكىست تاللاشتىكى بىر قانچە باسقۇچلار ئارقىلىق 10610 دانە جۈملە تاللىۋېلىنىپ بۇنۇڭ ئىچىدىكى 10000جۈملە مەشىقلەندۈرۈشكە قاتناشتۇرۇلدى،بۇ جۈملىلەرنىڭ كۇندۇلۇك تۇرمۇشتىكى قوللىنىلىش نىسپىتى بىر قەدەر يۇقىرى.سىناق جۈملىلىرى توپلام ئىچىدىكى سىناق جۈملىسى ۋە توپلام سىرتىدىكى سىناق جۈملىسى دەپ ئىككى تۈرگە ئايرىلغان بولۇپ ئايرىم-ئايرىم ھالدا 500 جۈملىدىن تەركىپ تاپتى.بۇ سانلىق مەلۇماتلار ئۆزى تەۋە بولغان توپلامدىن ئىختىيارى تاللىۋېلىندى.

سۆز تۈركۈمىنى بەلگىلەش نېگىزلىك ئېيتقاندا كۆپ خىللىق مەسلىسى （disambiguation） نى ھەل قىلىش بولۇپ ھازىرقى ئامبار(Alfa-ئامبار)دا جۈملە مۇھىتىدىكى ھەر بىر سۆزگە بىردىن سۆز تۈركۈمى بەلگىلەندى،جەمئى 123678 سۆزدىن تەركىپ تاپقان بولۇپ بۇنىڭ ئىچىدىكى 50.39% سۆز تۈركۈم جەھەتتە ھەرخىللىقتا ئىگە.2-جەدۋەلدە Alfa-ئامبىرىدىكى سۆزلەرنىڭ ھەرخىللىق ئەھۋالى ۋە مۇناسۋەتلىك سانلىق مەلۇماتلار كۆرسىتىلدى.

2-جەدۋەل Alfa-ئامبىرىدىكى سۆزلەرنىڭ ھەرخىللىق ئەھۋالى ۋە مۇناسۋەتلىك سانلىق مەلۇماتلار

ھەرخىللىق ئەھۋالى	تەكرارلانمىغان سۆزنىڭ ئومۇمى سانى
بىرلا خىللىق(بىرلا تۈركۈمگە تەۋە بولغىنى)	25412
ھەرخىللىقى(2-7 خىلغىچە تۈركۈمى بولغىنى)	3183
ئىككى تۈركۈمگە تەۋە بولغىنى	2497
ئۈچ تۈركۈمگە تەۋە بولغىنى	468
تۆت تۈركۈمگە تەۋە بولغىنى	165
بەش تۈركۈمگە تەۋە بولغىنى	52

يەتتە تۈركۈمگە تەۋە بولغىنى	1

بوغۇمنىڭ سىناق سانلىق مەلۇماتلارغا بولغان تەسىرى كۆزدە تۇتۇلۇپ تەجرىبە جەريانىدا بۇغۇم ئامىلىمۇ نەزەردىن ساقىت قىلىنمىدى،يەنى ئالدى بىلەن Alfa ئامبىرىغا قارتا بوغۇم ستاتىستىكىسى ئېلىپ بېرىلدى.مەزكۇر ئامباردا بىر سۆز كۆپ بولغاندىمۇ 10دانە بوغۇمدىن تەركىپ تاپقان بولۇپ،ئەڭ ئاز بولغاندا بىر بوغۇمدىن تەركىپ تاپقان.توققۇز ۋە ئون دانە بوغۇمدىن تەركىپ تاپقان سۆزدىن پەقەت بىر تالدىنلا بار،بۇغۇم ستاتىستىكا ئۇسۇلى ئامباردىكى سانلىق مەلۇماتلارنىڭ خاتالىقلارنى تۈزۈتۈپ ساپلىقىنى ئاشۇرۇشتا مۇھىم رول ئوينىدى.

3.2 تەجرىبە سۇپىسى ۋە نەتىجە ئانالىزى

تەجرىبە جەريانىدا مىكرو-سوفت شىركىتى ئوتتۇرغا قويغان ئەڭ يېڭى پىروگراممىا تۈزۈش تىلى بولغان C#،تېخنىكىلىق ئېچىش تىلى سۇپىتىدە قوللىنىلغان بولۇپ،ئۇ ئىشلىتىشكە ئەپلىك بولۇش،پۈتۈنلەي ئابونتقا يۈزلەنگەن بولۇش ۋە تۈر جەھەتتە بىخەتەر بولۇشتەك ئالاھىدىلىكلەرگە .بۇ ئالاھىدىلىكلەر ئۇنى كىينىكى بىر دەۋردىكى تەقسىمات تىپى قوللىنىشچان پىروگراممىسنىڭ ئاساسى ئېقىم ئىجادىيەت تىلىغا ئايلاندۇرغان[7]. .Net Framework 3.5 بولسا Linq تېخنىكىسىغا ئىگە بولۇپ Unicode بىر تەرەپ قىلىشقا ناھايىتى ماس كېلىدۇ. پىروگراممىا ئېچىش تىلى بولسا visual studio 2008 .steam system

ئېنتوناتسىيە قاتلاملىرىنى ئاپتۇماتىك بەلگىلەش تەتقىقاتى جەريانىدا-ئەتراپتىكى سۆزلەرنىڭ سۆز تۈركۈم ئالاھىدىلىكى،نۆۋەتتىكى سۆز ۋە كىينىكى سۆزدىكى جۈملىنىڭ ئۇزۇنلۇق ئالاھىدىلىكى،جۈملە بېشىدىن باشلاش ياكى جۈملە ئاخىرىدىن باشلاشتىكى ئالدىنقى چىگرا ئارلىق تىپى،ھەمدە ئەتراپىدىكى سۆزدە ئۇرغۇنىڭ بولۇش بولماسلىقى قاتارلىق ئالاھىدىلىكلەر بولسا نوقتىلىق ئويلىشىلىدىغان ئىلمىنتلار بولۇپ ھىساپلىنىدۇ[2]. سىناق باسقۇچىدا تەييارلىنىدىغان ئىككى خىل سىناق توپلىمىنىڭ بىرى مەشقىلەندۇرۇش توپلىمىدىن ئېلىنغان ئېچكى سىناق سانلىق مەلۇماتى،ئۇلار توپلامدىن ئىختىيارى تاللاپ ئېلىنغان،يەنە بىرسى بولسا مەشقىلەندۇرۇش توپلىمىنىڭ سىرتىدىن تاللىۋېلىنغان سىناق سانلىق مەلۇماتلاردۇر.سىستېمىنىڭ ئىجرا بولۇش كۆزنىكى تۆۋەندىكىچە:

4-رەسىم ئىجرا بولۇش

سىناقنىڭ توغرىلىق دەرىجىسىنى كۆزىتىشتە توغرىلىق نىسپىتى ۋە چاقىرىش نىسپىتى ھىساپلاندى.

توغرۇلۇق نىسپىتى=سۆز تۈركۈمى توغرا بەلگىلەنگەن سۆزنىڭ ئومۇمى سانى\سىناق تېكىست تەركىبىدىكى سۆز تۈركۈملىرىنىڭ ئومۇمى سانى

ھەر قايسى سۆز تۈركۈمىنىڭ چاقىرىش نىسپىتى=شۇ تۈركۈمگە تەۋە بولغان سۆزلەرنىڭ توغرا ئايرىغان ئۈمۈمى سانى\سىناق توپلامدىكى مەزكۇر سۆز تۈركۈمىنىڭ ئومۇمى سانى

چاقىرىش نىسپىتىنىڭ ئورتالانمىسى=ھەر قايسى سۆز تۈركۈملىرىنىڭ خاس چاقىرىش نىسپەتلىرىنىڭ يىغىندىسى\سۆز تۈركۈمىنىڭ سانى

3 -رەسىم سىناق نەتىجىسى

ئۇسۇل / باھالاش ئۆلچىمى	مەشقىلەندۈرۈش توپلىمىنىڭ ئىچىدىن تاللىۋېلىنغان سىناق سانلىق مەلۇماتلىرى 500 جۈملە	مەشقىلەندۈرۈش توپلىمىنىڭ سىرتىدىن تاللىۋېلىنغان سىناق سانلىق مەلۇماتلىرى 500 جۈملە
توغرۇلۇق نىسپىتى	86.92%	71.99%
چاقىرىش نىسپىتى	71.66%	62.42%

سۆز تۈركۈمى 41 خىلدىن 37 خىلغا ئازلىغاندا توپلام ئىچى سىنىقىنىڭ توغرۇلۇق نىسپىتى 86.69% دىن 86.92% غا ،چاقىرىش نىسپىتى 65.81% دىن 71.66% غا يۇقرى كۆتۈرۈلدى.توپلام سىرتى سىنىقىنىڭ توغرۇلۇق نىسپىتى بولسا 72.12% دىن 71.99% غا تۆۋەنلەپ ،چاقىرىش نىسپىتى 54.61% دىن 62.42% غا كۆتۈرۈلدى.گەرچە 37 خىل سۆز تۈركۈمى ئاساسىدىكى توپلام سىرتى سىنىقىنىڭ توغرۇلۇق دەرىجىسى سەل تۆۋەنلىگەن بولسىمۇ ئەمما باشقا قىممەتلەردە كۆرنەرلىك دەرىجىدە يۇقرى ئۆرلەش بولغان.تەجرىبە جەريانىدا ئېرىشىلگەن نەتىجە قوللىنىش تەلىپىگە ئۈيغۇن كېلىدۇ.شۇڭلاشقا 37 خىل سۆز تۈركۈمى ئاساسىدىكى ئىككى ئىلمىنىتلىق بولغان ئاپتۇماتىك سۆز تۈركۈمىنى بەلگىلەش ئالگورىتمى، ئېنتوناتسىيە قەۋەتلىرىنى ئاپتۇماتىك ئايرىش ئالگورىتمىنىڭ ئاساسى سۈپىتىدە قوللىنىلسا بولىدۇ.

4.ئاخىرلاشتۇرۇش سۆزى

بۇ ماقالىدە ئېنتوناتسىيە چىگرىسىنى ئاپتۇماتىك ئايرىش تېخنىكىسىغا يۆزلەندۈرۈلگەن ئۈيغۇر تىلىدىكى سۆز تۈركۈملىرىنى ئاپتوماتىك بەلگىلەش تېخنىكىسىغا قارتا تەتقىقات ئېلىپ بېرىلدى.بۇ قېتىملىق ئىزدىنىش ئارقىلىق،HMM مودىلى ئىككى ئېلىمىنتلىق ئۇسۇلىنىڭ،ئۈيغۇرتىلىدىكى ئېنتوناتسىيە چىگرىسىنى ئاپتۇماتىك ئايرىش جەريانىدا ئىشلىتىلىدىغان سۆز تۈركۈمىنى ئاپتۇماتىك بەلگىلەش تەلىپىگە ئۈيغۇن كېلىدىغانلىقى ئىسپاتلاندى.گەرچە بىر قىسىم خاتالىقلار مەۋجۇت بولسىمۇ ئەمما ئاساسى مەخسەت بولغان ئېنتوناتسىيە چىگرىسىنى ئاپتۇماتىك ئايرىش تەلىپىگە ئۈيغۇن ۋە ئېرىشىلگەن نەتىجە ئېھتىياجنى قاندۇرالايدۇ.ئەمما ھىساپلاش ئۇسۇلىغا نىسپەتەن يەنمۇ بىر قەدەم ئىلگىرىلىگەن ئاساستا ئىسلاھات ئېلىپ بېرىش كېرەك.ئۇندىن باشقا يەنە مەشقىلەندۈرۈلدىغان سانلىق مەلۇماتلار سانىنى ئاشۇرۇش،سۆز تۈركۈملىرىنىڭ سانى قاتارلىق مەسىلىلەردە يەنمۇ بىر قەدەم ئىلگىرىلىگەن ھالدا ئانالىز ئېلىپ بېرىش كېرەك.

参考文献

[1] 童方.语音合成系统的关键技术与应用实例[J].科海拾贝,2000.(2):19-21.

[2] Daniel Jurafsky,James H.Martin .Speech and Language Processing[M]. Beijing: Publishing house of Electronics Industry,2005.

[3] 艾斯卡尔·肉孜.基于HMM的维吾尔语音合成系统的研究与实现[D].新疆乌鲁木齐：新疆大学,2008.7.

[4] 洪梅,吐尔根·伊不拉音.维吾尔语的词性标注对校初谈[J].微型电脑引用,2006.(22):12.

[5] 田斌,易克初.一种可扩展的汉语统计二元文法语言模型[J].信号处理,2002.(3):184－187.

[6] Francis, W.N.and Kučera, H.(1982).Frequency Analysis of English Usage.Houghton,Boston.

[7] 康博.c#程序员参考手册[M].北京:清华大学出版社,2002.

[8] 赵晟,陶建华,蔡莲红.基于规则学习的韵律结构预测[J].中文信息学报,2002,16(5):30-37.

[9] 吴义坚.基于隐马尔科夫模型的语音合成技术研究[D].中国科学技术大学,2006.4.
[10] 杨行峻,迟惠生.语音信号数字处理[M].电子工业出版社,1995.
[11] 荀恩东,钱揖丽等.应用二叉树剪枝识别韵律短语边界[J].中文信息学报,2005.20(3):24-28.
[12] 陈永彬,王仁华.语音信号处理[M].中国科技大学出版社,1990.
[13] 吴志勇,蔡莲红.语音合成中的韵律关联模型[J].中文信息学报,2003,18(2):44-50.