Az ABSONIC beszédrögzítő, beszédkezelő és beszéd feldolgozó rendszer online beszédfelismerési szolgáltatása a GOOGLE és/vagy a NUANCE, együttesen mintegy 100 nyelvű illetve nyelvjárású speech-to-text előfizetésével (GOOGLE-2017.07.10. díjszabás: kb. 0.024 USD/perc ~ 400 Ft/óra ~ 400.000 Ft/1000 óra)

1. A beszédfelismerés nélküli, alap ABSONIC rendszer ismertetői: http://www.nicopro.hu/nicopro_v2/download/muszaki_ismerteto.pdf
2. Az ABSONIC beszédrögzítő, beszédkezelő és beszédfeldolgozó rendszer kiegészült egy online beszédfelismerési szolgáltatással, mellyel az élő beszéd valós időben (diktálás, értekezések, előadások, nyelvoktatás/tanulás, infokommunikációs beszélgetések is), míg a rögzített és letárolt beszéd utólag, automatikusan, megjeleníthető és/vagy karakteresen leíratható. Ennek integrált kivitele kvázi dobozos termék, amennyiben nincs igény az ABSONIC olyan speciális szolgáltatásaira, melyekhez tartozékok, kiegészítők és/vagy támogatás szükséges: http://www.nicopro.hu/nicopro_v2/index_hu.php?x=_HU/ismerteto_hu
3. A GOOGLE és a NUANCE beszédfelismertetése távoli, ismeretlen országokban történik, így a küldött beszédanyag és a kapott szöveganyag információ biztonsága nem olyan, mint a saját számítógépen illetve szerveren történő beszédfelismerésé, melyhez ***igény esetén, saját ABSONIC beszédfelismerővel, magyar esetleg angol nyelvű megoldást adunk, megfelelő előkészítés után (várhatóan ez heteket illetve hónapokat vehet igénybe).***
4. A beszédfelismerés pontossága a GOOGLE és/vagy a NUANCE beszédfelismerési pontosságától függ, tehát az ABSONIC csak kezelő felületek és a kapcsolódás szintjén a lehetőségét biztosítja a beszédfelismerésnek és annak pontosságát nem befolyásolja, a beszédfelismerés folyamatába nem avatkozik be. Emiatt az ABSONIC rendszert semminemű felelősség nem terheli az alábbi, fontosabb vonatkozásokban:
   4.1. minden a szolgáltatókhoz köthető ügy
   4.2. beszédfelismerés pontossága
   4.3. a regisztrációhoz köthető ügyek
   4.4. szolgáltatás minősége
   4.5. fizetési ügyek
   4.6. a szolgáltatás változásai (melyeket követni szándékozunk és a központi frissítés kezelési rendszerünkön keresztül gondozunk, amennyiben a szolgáltatók erre megfelelő lehetőséget biztosítanak)
   4.7. a GOOGLE rendszere még csak béta verzió, tehát még nem végleges, azaz lehetnek – várhatóan nem jelentős - változások, de már most is jól használható
5. A GOOGLE és/vagy a NUANCE beszédfelismerési pontossága az alábbiaktól függ, a teljesség igénye és lehetősége nélküli felsorolással (az alább felsoroltaknak egy része, bizonyos fokig kompenzálva van a következők tekintetében: zajszűrés, szintkiegyenlítés, visszhang mentesítés, stb.) :
   5.1. a beszélőnek a beszédkészsége (hangerő, hangminőség, beszédhibák, stb.)
   5.2. a beszélő nyelvismerete
   5.3. a beszélő környezetének akusztikai viszonyai, zajossága
   5.4. a beszédérzékelő eszköz minősége: mikrofon és/vagy infokommunikációs beszélőkészlet
   5.5. a beszélő és a beszédérzékelő eszköz (mikrofon, infokommunikációs beszélőkészlet, stb.) távolsága
   5.6. a beszédérzékelő eszköz beszédtovábbító rendszerének minősége (beszédátviteli eszközök, csatornák, stb.): mikrofonhálózatok, infokommunikációs hálózatok, adattovábbító és feldolgozó egységek, stb.
   5.7. a beszédfelismeréshez használt internet kapcsolat minősége
   5.8. a GOOGLE és a NUANCE együttesen több, mint 100 nyelvének illetve nyelvjárásának a beszédfelismerési kidolgozottsága, pontossága egymástól eltérő lehet. A két szolgáltató különféle nyelvjárásának egy része átfedésben van, azaz mindkettőnél megtalálható, míg a többi nyelv/nyelvjárás megoszlik a két szolgáltató között, tehát vagy az egyiknél van meg vagy a másiknál.

5.9. Igény esetén az előzőkben felsorolt, elsősorban helyszínfüggő, negatív tényezők közül néhányat akár jelentősen is enyhíteni, befolyásolni tudunk, előzetes egyeztetés, helyszíni szemle és külön megállapodás keretében.

6. Az élő beszéd valós időben is felismertethető bizonyos késleltetéssel (hozzávetőleges becsléssel: általában 2-10 esetleg több másodperccel), mely előre nem határozható meg (ehhez helyszínföggő, felhasználói próbák kellenek):
   6.1. diktálások azonnali megjelenítése és leíratása
   6.2. értekezletek azonnali megjelenítése és leíratása
   6.3. előadások valós idejű megjelenítése az előadótermek kivetítőin és a hallgatóság számítógépein
   6.4. nyelvoktatás és nyelvtanulás
   6.5. infokommunikációs beszélgetések azonnali megjelenítése és leíratása
7. A letárolt beszéd leíratása esetén nincs olyan valós idejű megjelenítési igény, mint az élő beszédnél, így ennek leíratási időszükséglete alapvetően a letárolt beszéd időtartamától függ.
8. A felismertetett szöveg az ABSONIC-ba beépített szövegszerkesztőben szerkeszthető, RTF formátumban
9. A letárolt beszéd és a felismertetett szöveg kezelhető együtt és külön-külön is (szerkesztés, export-import, archiválás stb.).
10. A beszédfelismerés eredményeként létrejövő szöveg a GOOGLE előfizetéses, online fordítóprogramjának átadható lesz, amint ezt a fejlesztési modult is befejezzük.
11. Az ABSONIC rendszer jelenleg egy időben csak egy beszédfelismerést tesz lehetővé, annak ellenére, hogy a rendszer többcsatornás. Az egyidejűleg több csatornán, különböző nyelveken történő beszédfelismerés fejlesztés alatt van. Ez jelenthet majd több mikrofon csatornát (például tolmácsok) vagy tízes, százas nagyságrendű infokommunikációs csatornát is: analóg telefon, ISDN, VoIP, GSM, Viber, WhatsApp, Messenger, műholdas telefonok-INMARSAT-IRIDIUM-THURAYA, stb.
12. A GOOGLE mindenkor aktuális nyelv és nyelvjárás készlete a következő linken található meg (a jelenlegiek mellékelve: 2017.07.10.), így ez alapján kipróbálhatja és eldöntheti, hogy melyik szolgáltatót mikor, mire használja: https://cloud.google.com/speech/docs/languages
   12.1. az előfizetéshez illetve regisztrációhoz GOOGLE fiókkal kell rendelkezni
   12.2. a szolgáltatóval való kapcsolatkezeléshez, így a regisztrációhoz is megfelelő angol tudás szükséges és megfelelő informatikai gyakorlat illetve ismeretek (WEB-es vonatkozások). A szolgáltatóhoz való kapcsolódásból fakadó mindennemű tevékenység, ügylet, esemény, jelenség a regisztrált felhasználó kizárólagos és teljeskörű felelőssége minden vonatkozásban.
   12.3. az előfizetés regisztrációhoz kötött egy bankkártya adatainak megadása mellett: https://console.cloud.google.com/ - jobb felső sarkokban Sign up for free trial - Ország megadása, majd mindkét kérdésre igen válasz - Következő lapon meg kell adni a cég és/vagy személy adatokat, bankkártya adatokat.
   12.4. fizetés a felhasznált időtartam szerint, időszakonként (havonta), bankkártyáról történő leemeléssel, az ingyenes keretek felhasználása után
   12.5. havonta 0-60 perc beszédfelismerés ingyenes
   12.6. havonta a 60. perctől a díj: 0.006 USD/15 másodperc = 0.024 USD/perc ~ 400 Ft/óra ~ 400.000 Ft/1000 óra
   12.7. első alkalommal, az első évre szólóan minden regisztrált 300 USD ~ 12.500 perc ~ 200 óra ingyenes keretet kap, amely az első évben, egy évig felhasználható
   12.8. A Google díjszabása megváltozhat, így rendszeresen ellenőrizze azt!
13. A NUANCE mindenkor aktuális nyelv és nyelvjárás készlete a következő linken található meg (a jelenlegi mellékelve: 2017.07.10.), így ez alapján kipróbálhatja és eldöntheti, hogy melyik szolgáltatót mikor, mire használja: https://developer.nuance.com/public/index.php?task=supportedLanguages
   13.1. az előfizetés regisztrációhoz kötött: https://developer.nuance.com/public/index.php?task=prodStart

13.2. a szolgáltatóval való kapcsolatkezeléshez, így a regisztrációhoz is megfelelő angol tudás szükséges és megfelelő informatikai gyakorlat illetve ismeretek (WEB-es vonatkozások). A szolgáltatóhoz való kapcsolódásból fakadó mindennemű tevékenység, ügylet, esemény, jelenség a regisztrált felhasználó kizárólagos és teljeskörű felelőssége minden vonatkozásban.

13.3. fizetés PayPal-en és a hozzá kapcsolható bankkártyán illetve bankszámlán keresztül (email cím és jelszó), előre fizetéssel történik, az előre fizetési minimális összege 25USD, ami 3.125 tranzakció illetve mondat. A későbbi fizetések alkalmankénti, személyes közreműködéssel vagy automatikus leemeléssel is történhetnek (direct debit), országonként esetleg eltérően.

13.4. 20.000 tranzakció, azaz mondat felismerése havonta ingyenes, e-felett 0.008 USD/tranzakció illetve mondat. A beszédben tartott szünetek alapján osztja fel a szöveget mondatokra a rendszer. Hozzávetőleges, statisztikai becslés alapján egy mondat maximálisan 30 másodperc, mivel ennyi ideig tud egy ember egyfolytában beszélni levegővétel nélkül, de valójában 10 másodperc egy mondat átlagos ideje. Ezekkel a becslésekkel a 20.000 tranzakció/mondat kb. 10.000 - 30.000 perc ~ 167 – 500 óra ingyenes, havonta. Megjegyzés: Amennyiben olyan gyors a beszéd, hogy a mondatok közötti szünetek nem érzékelhetők megfelelően illetve egyéb okból a mondatok nem határolhatók be pontosan, úgy a tranzakciókat a hosszuk alapján határozzák meg, melynél egy tranzakció maximális hossza 2 Mbyte, ami kb. megfelel 1 (egy) percnyi 16 bit-es, 16 kHz-es, PCM mono audio jelnek illetve beszédnek és ekkor ilyen adatcsomagokban történik a beszédfelismerés és a költségelése.

13.5. lehetőség van a szótár utántöltésére a beszédfelismerő rendszer által nem ismert szavak tekintetében, a feltöltés file-okkal történhet (WORD, TXT, stb): https://developer.nuance.com/public/index.php?task=uploadVocabulary

13.6. A NUANCE díjszabása megváltozhat, így rendszeresen ellenőrizze azt!

14. Amennyiben az előzőekhez kérdése, javaslata, véleménye van, úgy keressen bennünket: nicopro@nicorpo.hu

GOOGLE and NUANCE languages-2017-07-10				
No	**Country**	**Languages**	Google	Nuance
1.	Algeria	Arabic	x	x
2.	Argentina	Spanish	x	x
3.	Australia	English	x	x
4.	Austria	German		x
5.	Bahrain	Arabic	x	x
6.	Bangladesh	Bengal		x
7.	Belgium	Dutch		x
8.	Belgium	French		x
9.	Bolivia	Spanish	x	
10.	Brazil	Portuguese	x	x
11.	Bulgaria	Bulgarian	x	x
12.	Canada	English	x	x
13.	Canada	French	x	x
14.	Canada	Hindi		x
15.	Chile	Spanish	x	
16.	China	Chinese, Mandarin (Simplified)	x	x
17.	Colombia	Spanish	x	x
18.	Costa Rica	Spanish	x	
19.	Croatia	Croatian	x	x
20.	Czech Republic	Czech	x	x
21.	Denmark	Danish	x	x
22.	Dominican Republic	Spanish	x	
23.	Ecuador	Spanish	x	
24.	Egypt	Arabic	x	x
25.	El Salvador	Spanish	x	
26.	Finland	Finnish	x	x
27.	France	French	x	x
28.	Germany	German	x	x
29.	Greece	Greek	x	x
30.	Guatemala	Spanish	x	
31.	Honduras	Spanish	x	
32.	Hong Kong	Chinese, Cantonese (Traditional)		x
33.	Hong Kong	Chinese, Mandarin (Simplified)	x	x
34.	Hungary	Hungarian	x	x
35.	Iceland	Icelandic	x	
36.	India	English	x	x
37.	India	Assamese		x
38.	India	Bhojpuri		x
39.	India	Bengali		x
40.	India	Gujarati		x
41.	India	Hindi	x	x
42.	India	Marathy		x
43.	India	Oryja		x
44.	India	Punjabi		x
45.	India	Telugu		x

No	Country	Languages	Google	Nuance
46.	India	Urdu		x
47.	India, Malayzia	Tamil		x
48.	Indonesia	Bahasa	x	x
49.	Indonesia	Indonesian		
50.	International	Arabic		x
51.	Iran	Persian		x
52.	Iraq	Arabic	x	x
53.	Ireland	English	x	
54.	Israel	Arabic	x	
55.	Israel	Hebrew	x	x
56.	Italy	Italian	x	x
57.	Japan	Japanese	x	x
58.	Jordan	Arabic	x	
59.	Kuwait	Arabic	x	x
60.	Latin-America	Spanish		x
61.	Lebanon	Arabic	x	x
62.	Lithuania	Lithuanian	x	
63.	Malaysia	Malay	x	x
64.	Malayzia	Hindi		x
65.	Mexico	Spanish	x	x
66.	Morocco	Arabic	x	x
67.	Nepal	Nepali		
68.	Netherlands	Dutch	x	x
69.	New Zealand	English	x	
70.	Nicaragua	Spanish	x	
71.	Norway	Norwegian Bokmål	x	x
72.	Oman	Arabic	x	x
73.	Pakistan	Urdu		
74.	Panama	Spanish	x	
75.	Paraguay	Spanish	x	
76.	Peru	Spanish	x	x
77.	Philippines	English	x	
78.	Philippines	Filipino	x	
79.	Poland	Polish	x	x
80.	Portugal	Portuguese	x	x
81.	Puerto Rico	Spanish	x	
82.	Qatar	Arabic	x	x
83.	Romania	Romanian	x	x
84.	Russia	Russian	x	x
85.	Saudi Arabia	Arabic	x	x
86.	Serbia	Serbian	x	x
87.	Singapore	English		
88.	Slovakia	Slovak	x	x
89.	Slovenia	Slovenian	x	
90.	South Africa	Afrikaans	x	
91.	South Africa	English	x	
92.	South Africa	Zulu	x	
93.	South-Korea	Korean		x
94.	Spain	Basque	x	

No	Country	Languages	Google	Nuance
95.	Spain	Catalan	x	x
96.	Spain	Galician	x	x
97.	Spain	Spanish	x	x
98.	Spain	Valencian		
99.	State of Palestine	Arabic	x	x
100.	Sweden	Swedish	x	x
101.	Switzerland	German		x
102.	Taiwan	Chinese, Mandarin (Traditional)	x	x
103.	Thailand	Thai	x	x
104.	Tunisia	Arabic	x	x
105.	Turkey	Turkish	x	
106.	Ukraine	Ukrainian	x	x
107.	United Arab Emirates	Arabic	x	x
108.	United Kingdom	English	x	x
109.	United States	English	x	x
110.	United States	Spanish	x	x
111.	Uruguay	Spanish	x	
112.	Venezuela	Spanish	x	
113.	Vietnam	Vietnamese	x	x
114.	Wales	Welsh		x

DEV

Google's speech recognition technology now has a 4.9% word error rate

EMIL PROTALINSKI @EPRO MAY 17, 2017 4:06 PM

Google CEO Sundar Pichai today announced that the company's speech recognition technology has now achieved a 4.9 percent word error rate. Put another way, Google transcribes every 20th word incorrectly. That's a big improvement from the 23 percent the company saw in 2013 and the 8 percent it shared two years ago at I/O 2015.

The tidbit was revealed at Google's I/O 2017 developer conference, where a big emphasis is on artificial intelligence. Deep learning, a type of AI, is used to achieve accurate image recognition and speech recognition. The method involves ingesting lots of data to train systems called neural networks, and then feeding new data to those systems in an attempt to make predictions.

"We've been using voice as an input across many of our products," Pichai said onstage. "That's because computers are getting much better at understanding speech. We have had significant breakthroughs, but the pace even since last year has been pretty amazing to see. Our word error rate continues to improve even in very noisy environments. This is why if you speak to Google on your phone or Google Home, we can pick up your voice accurately."

For the sake of comparison, Microsoft declared in October 2016 that it had reached speech recognition parity with humans. Its word error rate at the time was 5.9 percent, though it's not clear if the two companies are following the same standards of evaluation.

Google has been touting its speech recognition improvements for a while now. Earlier this year, the company said it had slashed its speech recognition word error rate by more than 30 percent since 2012. The main reason for the drastic improvement? Google confirmed that it's the use of neural networks.

Pichai also shared an interesting tidbit about Home's development: "When we were shipping Google Home, we were originally planning to include eight microphones... But thanks to neural networks, using a technique called 'neural beam forming', we were able to ship it with just two microphones and achieve the same quality."

So if you're surprised at how well (or poorly) Google understands what you're saying, this is why. Recognition is getting better and better, but there's still room to get that word error rate closer to 0 percent.

Google I/O 2017: Get the latest news here

VIDEO

Automatic Speech Recognition Services Comparison

Automatic Speech Recognition Services Comparison

Introduction

“Ok Google, find me a red dress.” Your long-time customer has just been invited to an important party this evening and wants to make a good impression. She’s on her way to your store right now and can’t spend any time typing in searches while she drives. Instead of saying, “Ok, Google…” wouldn’t you rather she said, “Ok, MyFavoriteStore name?”

Both Apple and Google have done a good job educating users on the value and ease of voice-controlled features. So how mature is commercial speech recognition today? As Grid Dynamics has extensive experience in eCommerce and search solutions, we decided to take a look at the current speech recognition technologies available for voice search implementation. In this article we will share the results from our experiment - comparing the quality of different speech recognition providers.

Services

Before the Experiment was started, our team reviewed multiple providers of automatic speech recognition. We have used the following criteria for selection of the service to evaluate:

- Unified, cross-platform interface. It means service availability via HTTP REST interface
- Speech recognition quality “out of the box” without any tuning for particular customer
- Free (or low price) for initial testing of service
- Speech recognition provided as a SaaS [https://en.wikipedia.org/wiki/Software_as_a_service]

We compared the following services.

- Google
- Nuance [http://www.nuance.com/index.htm]
- AT&T [https://developer.att.com/]

- WIT [https://wit.ai/]
- IBM Watson [http://www.ibm.com/smarterplanet/us/en/ibmwatson/developercloud/]

Block Join Faceti... 2
Block Join Faceting: Intr...
Block Join Faceting in S...
Automatic Speec... 9
Scoring Join Part... 3
How to import structured...
Lucene SIMD Codec be...
Who is who in Big Data
Spark and ZooKeeper: f...
Proposing SIMD ... 1

Numeric Range Queries...
Alternative approach to ...
Segmented Filter ... 13
Grandchildren an... 4
Solr block-join su... 49

Block Join Query ... 6
Ignoring test failures at CI
Highlights from our Ope...

Solr Experience: ... 10

Solr Experience: ... 4

Spring Nested - Part III

Google

Google Speech API is not “production” ready.

- Experimental status can change API at any time
- No official API documentation or usage capabilities
- Limitations of approximately 500 requests per day, per account
- You need to join Chromium-dev mail group [https://groups.google.com/a/chromium.org/forum/?fromgroups#!forum/chromium-dev] and generate appropriate key in Google developer console [https://console.cloud.google.com/home/dashboard]

Example of API usage:

```
curl -X POST \
--header 'Content-Type: audio/x-flac; rate=44100;' \
--data-binary @red_dress.flac \
'https://www.google.com/speech-api/v2/recognize?lang=en-us&key=<KEY>'
```

Nuance

Nuance speech recognition REST API features:

- Registration [https://developer.nuance.com/public/index.php?task=register] is required
- Account upgrade from Silver- to Gold-level offered for free
- Usage limitations of 5,000 requests per day

Example of API usage:

```
curl -X POST \
--header "Content-Type: audio/x-wav;codec=pcm;bit=16;rate=16000" \
--header "Accept: application/xml" \
--header "Accept-Topic: Dictation" \
--data-binary @red_dress.wav \
"https://dictation.nuancemobility.net:443/NMDPAsrCmdServlet/dictation?appId=<APP_ID>&appKey=<APP_KEY>"
```

AT&T

AT&T speech recognition REST API features:

Block Join Faceti... 2
Block Join Faceting: Intr...
Block Join Faceting in S...
Automatic Speec... 9
Scoring Join Part... 3
How to import structured...
Lucene SIMD Codec be...
Who is who in Big Data
Spark and ZooKeeper: f...
Proposing SIMD ... 1
Numeric Range Queries...
Alternative approach to ...
Segmented Filter ... 13
Grandchildren an... 4
Solr block-join su... 49
Block Join Query ... 6
Ignoring test failures at CI
Highlights from our Ope...
Solr Experience: ... 10
Solr Experience: ... 4
Spring Nested - Part III

- Registration [https://developer.att.com/developer/flow/apiPlaygroundFlow.do?execution=e7s1] is required
- Required “Premium Access” payment is $99/year + Usage fees [http://developer.att.com/pricing/speech-pricing-details] to access automatic speech recognition
- AT&T REST API [http://developer.att.com/apis/speech/docs] uses OAuth 2.0 for authorization
- According to documentation usage limitations is 1 request per second [https://developer.att.com/support/faqs/att-developer-program-and-api-platform-faqs#what-are-maximum-transaction-rates-for-apis]

Example of API usage:

```
curl -X POST \
--header "Authorization: Bearer <TOKEN>" \
--header "Content-Type: audio/x-wav"   \
--data-binary "@red_dress.wav"   \
"https://api.att.com/speech/v3/speechToText"
```

WIT

WIT is more about NLP [https://en.wikipedia.org/wiki/Natural_language_processing] (Natural Language Processing) than about plain-speech recognition.

- Main focus, besides speech recognition, is to parse out spoken phrases and extract valuable information (e.g., some voice command). The goal is to have the system “understand” voice. For example, play “Jingle Bells” when the user says, “Hi, robot! Please play me Christmas songs.”
- Github account is all that is needed to access WIT REST API [https://wit.ai/docs/http/20141022]
- No account usage limitation

Example of API usage:

```
curl -X POST \
--header "Authorization: Bearer <TOKEN>" \
--header "Content-Type: audio/wav" \
--data-binary "@red_dress.wav"   \
"https://api.wit.ai/speech?v=20141022"
```

IBM Watson

IBM Speech recognition REST API features:

- Public API was released to the public in early 2015

- Registration in Bluemix [https://console.ng.bluemix.net/] is required
- Usage limitations of 150,000 requests per month

Example of API usage:

```
curl -X POST \
--header "Content-Type: audio/flac" \
--user <USERNAME>:<PASSWORD> \
--data-binary "@red_dress.flac"  \
"https://stream.watsonplatform.net/speech-to-text/api/v1/recognize"
```

Experiment

To compare quality of speech recognition, you first need a recorded voice. As we worked on voice search features for eCommerce, we recorded eCommerce-like search phrases. We used short phrases such as: brand names, colors, sizes, etc. Here's a sample of the phases used - "red dress," "Calvin Klein jeans" and "xl coat." We leveraged over 3,000 different phrases for this experiment and compared different conditions like gender, age and background noise (I.e., with or without noise), as well as other criteria.

We used the following sequence for experiment purposes.

- Delivered an audio file with recorded search phrases to external services
- Received recognized text from automatic speech recognition service
- Evaluated quality metrics of recognized text vs. actual search phrase

We used multiple quality metrics, such as:

- Volume of exact recognized phrases
  - Simple, but a paramount quality metric
  - Larger number of exact recognized phrases, the better quality of speech recognition results
- Word Error Rate [https://en.wikipedia.org/wiki/Word_error_rate] (WER)
  - Minimum number of words edits (I.e., insertions, deletions or substitutions) required to change one phrase into the other
  - Normalized by phrase length (basically leveraging Levenshtein distance [https://en.wikipedia.org/wiki/Levenshtein_distance] between two phrases working at the word level, instead of the phenomenal level)
  - Fewer number of required edits, which meant that the phrases are more like each other - offering the best quality of speech recognition

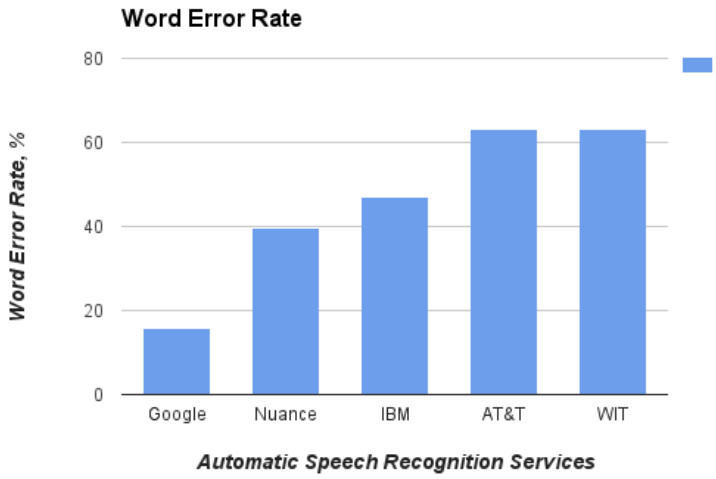
Block Join Faceti... 2
Block Join Faceting: Intr...
Block Join Faceting in S...
Automatic Speec... 9
Scoring Join Part... 3
How to import structured...
Lucene SIMD Codec be...
Who is who in Big Data
Spark and ZooKeeper: f...
Proposing SIMD ... 1
Numeric Range Queries...
Alternative approach to ...
Segmented Filter ... 13
Grandchildren an... 4
Solr block-join su... 49
Block Join Query ... 6
Ignoring test failures at CI
Highlights from our Ope...
Solr Experience: ... 10
Solr Experience: ... 4
Spring Nested - Part III

Comparison Results

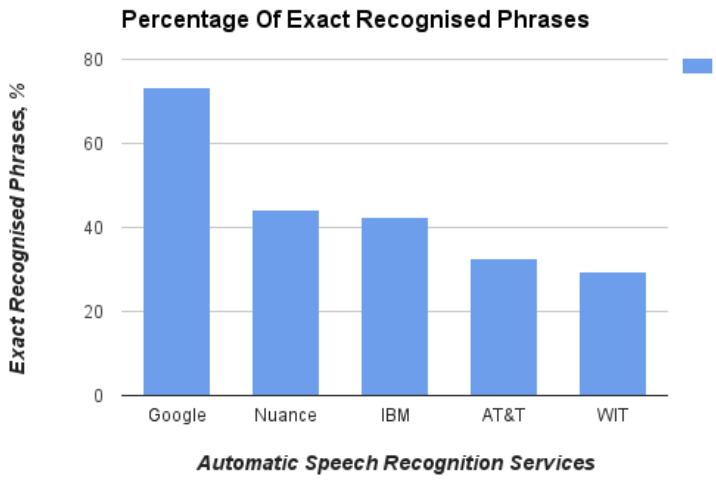
The quality champion is Google. We didn't reproduce the declared by Google 8% WER [http://venturebeat.com/2015/05/28/google-says-its-speech-recognition-technology-now-has-only-an-8-word-error-rate/] with our Grid Dynamics' data, but the results are still impressive. Google achieved 73.3% of exact recognized phrases with a 15.8% WER.

Nuance came in second place by a large margin. In Nuance, 44.1% of the phrases were recognized perfectly and the WER was 39.7%. IBM (46.9.3% and 42.3% WER) came in third place. While AT&T and WIT had the exact same WER - 63.3%, with a small advantage in exact recognition by AT&T (32.8% vs 29.5%, WIT).

Word Error Rate (less is better):



Percentage of Exact Recognized Phrases (more is better):



Block Join Faceti... 2
Block Join Faceting: Intr...
Block Join Faceting in S...
Automatic Speec... 9
Scoring Join Part... 3
How to import structured...
Lucene SIMD Codec be...
Who is who in Big Data
Spark and ZooKeeper: f...
Proposing SIMD ... 1
Numeric Range Queries...
Alternative approach to ...
Segmented Filter ... 13
Grandchildren an... 4
Solr block-join su... 49
Block Join Query ... 6
Ignoring test failures at CI
Highlights from our Ope...
Solr Experience: ... 10
Solr Experience: ... 4
Spring Nested - Part III

Conclusion

Based on our test criteria of exact recognized phrases and word error rate, Google is by far the best solution out of the box. This is not surprising given their history of developing and proving voice search, but unfortunately - for now - it is not commercially available. Google's quality, however, could be used as a benchmark for the commercially available products as many of them have tools and features for customizing search experience.

Exact phrase match and word error rate are only two issues to provide world-class voice search that your customers will soon expect. Additional challenges are speech recognition performance and recognizing eCommerce-specific terms. For instance, consider searches like brands, sizes, materials and, of course, long/complex phrase recognition (I.e., "Ok, MyFavoriteRetailer, find me a Ralph Lauren or Ann Taylor red cocktail dress, knee length and open back, in a size 9 that isn't dry-clean only").

But, we will discuss those challenges and our solutions in future articles.

Posted 11th January 2016 by Andrey Kudryavtsev

Labels: API, speech recognition, voice search, ~Andrey Kudryavtsev

Block Join Faceti... 2
Block Join Faceting: Intr...
Block Join Faceting in S...
Automatic Speec... 9
Scoring Join Part... 3
How to import structured...
Lucene SIMD Codec be...
Who is who in Big Data
Spark and ZooKeeper: f...
Proposing SIMD ... 1
Numeric Range Queries...
Alternative approach to ...
Segmented Filter ... 13
Grandchildren an... 4
Solr block-join su... 49
Block Join Query ... 6
Ignoring test failures at CI
Highlights from our Ope...
Solr Experience: ... 10
Solr Experience: ... 4
Spring Nested - Part III