

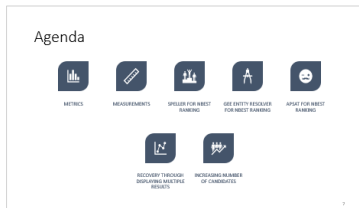


SR Recovery

Speech Continuation

N-Best Rankers and Oracle for Bing Search

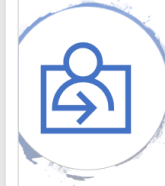
Slides ahead



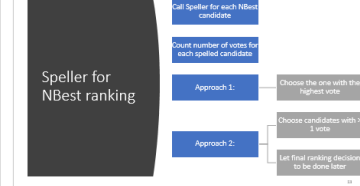
Metrics

- Metrics**
 - WER (Levenshtein)
 - SIM (Similarity)
 - Mean-CI (statistical impact on user understanding)
 - It is likely to happen when WER per word in a query $\Rightarrow 0.5$
 - Accuracy – absolute match with transcribed query
- Data**
 - 1200 samples, Cortana traffic
 - 1200 samples, Bing traffic
 - Transcribed by vendors, corrected by FTE

-1K	SIM	WER	CI (%)	Accuracy
	0.9776	1.1009	9.74	83%



- ## Terms
- SR** = Speech recognition
 - SR candidate (SRc)** = candidates provided by SR
 - SRc Weight** = suggested as best
 - Speller suggestion** = first suggestion that Speller provides for an SR candidate. Speller also may have own candidates for suggestion, with different level confidence
 - Speller Ranker** = one based on patterns of Speller suggestions
 - NBcSR Ranker** = one based on NBcSR Order from SR
 - Vote** = scores assigned by Speller Ranker to an SR candidate
 - "Misplaced"** = not a traditional notion that the utterance is misparsed in the sense of traditional English Grammar. Rather, Speller SR returns a use alternative (so called "suggestion").



Speller and NBest0 Rankers agree

id	id candidate	elect correct	reinstatement	seller suggestion	research query	vote
0	where is the nearest public bathroom	1	3	where is the nearest public bathroom	where is the nearest public bathroom	0.5
1	where the nearest public bathroom	3	3	where is the nearest public bathroom	where is the nearest public bathroom	0
2	where is the nearest public bathroom	1	3	where is the nearest public bathroom	where is the nearest public bathroom	0
3	where is the nearest public bathroom	3	3	where is the nearest public bathroom	where is the nearest public bathroom	0.5
4	where is the nearest public bathroom	1	3	where is the nearest public bathroom	where is the nearest public bathroom	0.5

Speller Ranker wins (based on CI by 1.94% based on accuracy 2.12%)

id	id candidate	classcorrect	rankcorrect	spliter suggestion	Transcribed query	Vote
0	element paper john's pizza	0	1	element paper john's pizza	element paper john's pizza	0
1	element paper john's pizza	0	1	element paper john's pizza	element paper john's pizza	0
2	element paper john's pizza	0	1	element paper john's pizza	element paper john's pizza	0
3	element paper john's pizza	0	1	element paper john's pizza	element paper john's pizza	0
4	element paper john's p	0	1	element paper john's p	element paper john's pizza	0

NBest0 Ranker wins

Id	Set candidate	document	retrieval	qualifier suggestion	Transcribed query	Size
1	find me a bang and shafuku tv	1	1	1 find me a bang and shafuku tv	find me a bang and shafuku tv	0.0
2	find me a bang shafuku tv	1	1	1 find me a bang shafuku tv	find me a bang and shafuku tv	0.0
3	find me a bang shafuku tv	1	1	1 find me a bang shafuku tv	find me a bang and shafuku tv	0.0
4	find me a bang shafuku tv	1	1	1 find me a bang shafuku tv	find me a bang and shafuku tv	0.0
5	find me a bang shafuku tv	1	1	1 find me a bang shafuku tv	find me a bang and shafuku tv	0.0

Recovery: Entity Resolver: Bang/Shafuku TV

Recovery: Entity Resolver: Bang&Olufsen TV

Speller and NBest0 are both wrong

Id	IR candidate	character	character	Similar suggestion	Character query	Size
0	what is the vesper wending club		0	0what is the vesper wending club	what is the vesper rolling club	
1	what is the vesper rolling club		0	0what is the vesper rolling club	what is the vesper rolling club	
2	what is the vesper rowing club		0	0what is the vesper rowing club	what is the vesper rolling club	
3	what does the vesper rowing club		0	0what does the vesper rowing club	what is the vesper rolling club	
4	what is the vesper wending club		0	0what is the vesper wending club	what is the vesper rolling club	

Recovery: Entity Resolver: Vesper Rolling Club not found

Speller and NBest0 are both wrong, revisited

id	SA candidate	thetacorrect	thetaincorrect	Quoter suggestion	thetacorrect query	date
1	what is the vesper meeting club	0		What is the vesper meeting club	what is the vesper meeting club	
2	what is the vesper rolling club	0		What is the vesper rolling club	what is the vesper rolling club	
3	what is the vesper rowing club	0		What is the vesper rowing club	what is the vesper rowing club	
4	what does the vesper rowing club	0		What does the vesper rowing club	what is the vesper rowing club	
5	what is the vesper meeting club	0		What is the vesper meeting club	what is the vesper meeting club	

Recovery: Entity Resolver: Vesper Rolling Club... also likely a transcriber's error: **Vesper Rowing Club**
Reasons: There is a candidate that is identical to the second Speller suggestion, with a competitive regression score

GEE entity
resolver for
NBest ranking

Look up entities for (remaining) NBest candidates

Choose the longest candidates(s) with most words as entities

Generic Entity Extraction for Queries: What is the goal?

- Identify all entities in a **query**. Return their **spans**, **repository IDs** and their **type**.



APSAT for
NBest ranking

Look up predicted APSAT for (remaining) NBest candidates

Choose the one with the highest value (if the difference is statistically significant).

What is APSAT?

Answer Prominence Satisfaction

Purpose: Measures answer satisfaction at Policy-Vote

A prominent answer must satisfy the majority (50%+) of users. If more than 50% of

Intuition answered as unhelped by another answer(s), or abandoned without satisfaction, then it's a DSAT for the prominent answer.

Assumptions We assume that, when there aren't any page clicks, the user has seen the prominent answer given its placement on the page, so abandonment can apply

Usage	Rankers predict AP SAT for every answer to determine answer placement on the page.
--------------	--

Answers ship criteria Existing answers - AFSAT > 0. New Answers - AFSAT Win Rate > 10%

Note: AFSAT only applies to assets at TDR. Surplus applies to everything else



Bing recovers

- "earls kitchen + bar"
- SR NBESTo "earls kitchen and bar"
- Speller: "earls kitchen and bar (all candidates)"
 - Speller is formally wrong
- Satori QAS: entity: "kitchen"
 - Doesn't help

- APSAT:

- Bing: right answer is in results as #2

Gaps



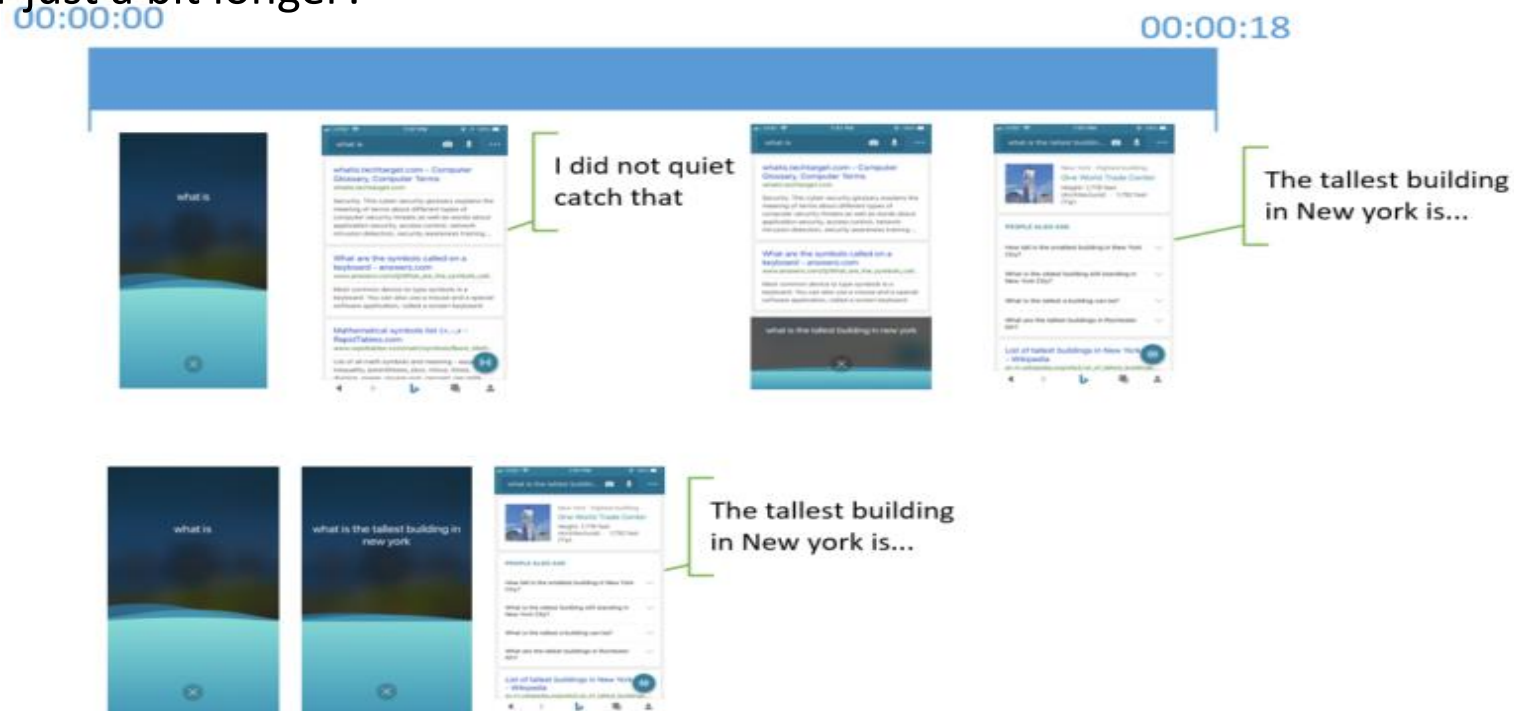
More
candidates
improve
recovery

- **<N> candidates for NBest**
 - Currently fixed to N=5
 - Potential to get ~25 per query from SR (up to 50)
 - Early experimentation results: N>5, in general, and should be optimized per record (e.g., Classifier N=foo(words in query), confidence, ranker scores)
- **<M> suggestions for Speller**
 - Up to 50
 - Use regression and lambda scores
- **(Additional) Homophones**

Speech continuation

Problem

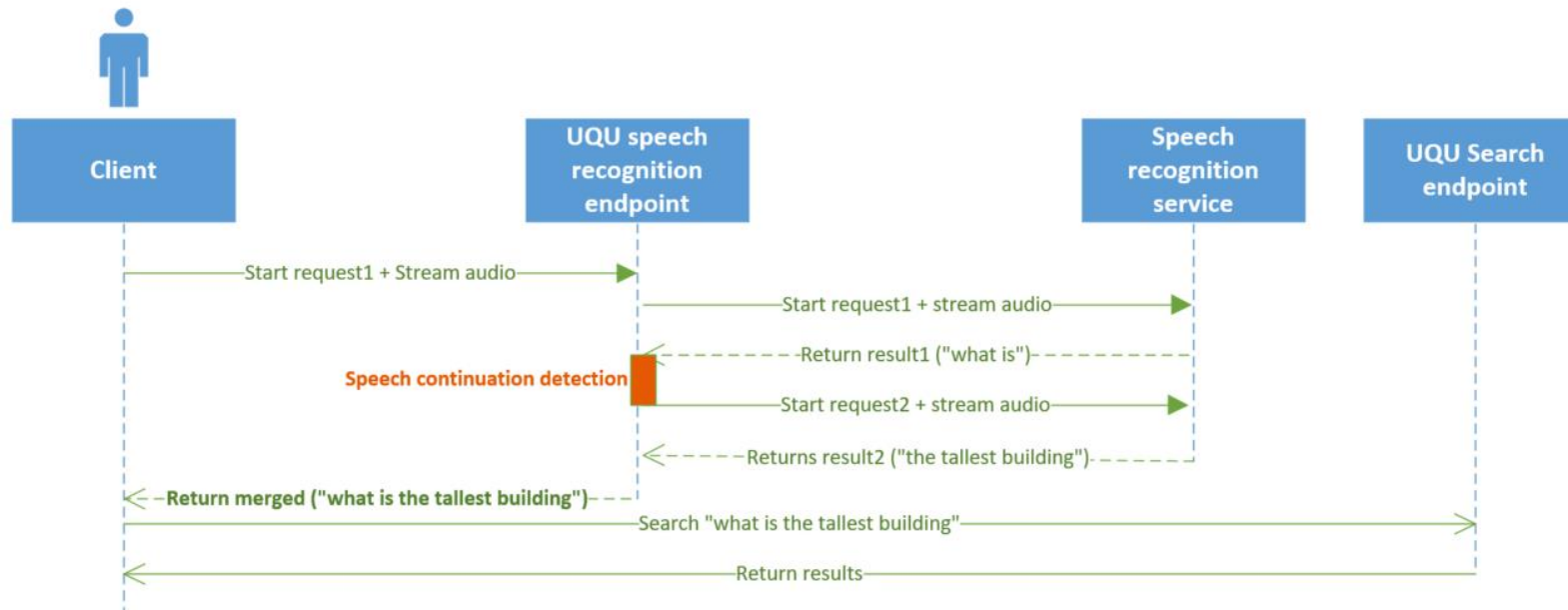
- About 1% of all speech queries triggered incomplete query skill in past 7 days
- 16% of these had a follow up turn which started with the same query.
- Recovery time from being cutoff mid speech can be over 10 seconds because user must wait for incomplete results to load, and then reopen mic and reissue query
- Speech stops recognizing audio after silence has been detected. What if we could keep mic open for just a bit longer?



Speech continuation

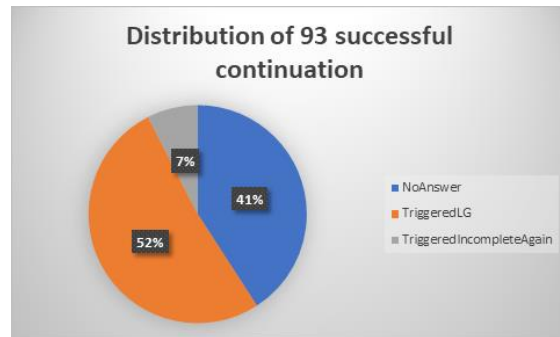
Solution

1. Detect user is likely to continue query
2. Intercept first result from SR service and do not send it to client
3. Start new recognition with SR service and continue streaming audio
4. Continue streaming intermediary results to client to update UI
5. Merge the new results from SR service with original first result



Speech continuation Results

- Current classifier is based on small set 33 incomplete phrases such as “what is”
 - Classifies about 0.25% of queries as incomplete
- 35% of users continued speaking when given the chance
 - 52% of those successfully triggered a Bing answer (compared to just blue links)



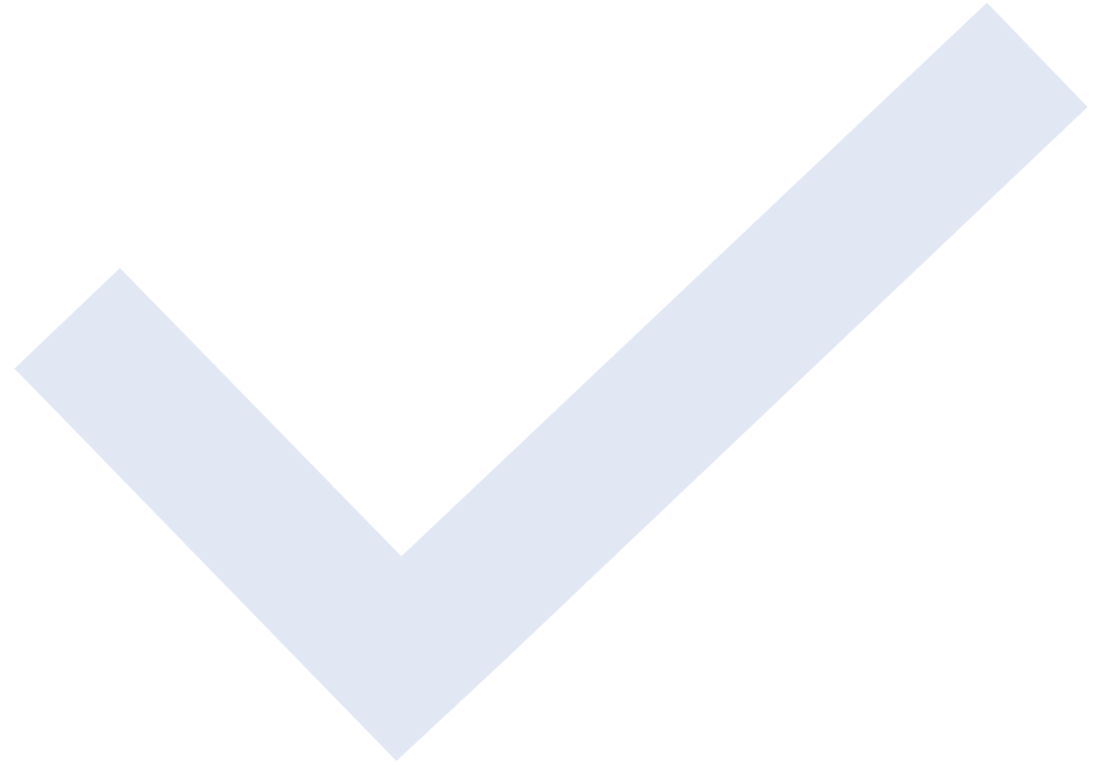
- Example continued queries:

query0	query1
what is the best	plant for shady area
2011 toyota prius	hybrid batteries
costco corporate offices	phone number
what is normal	cold water PH



SR Recovery

Lessons of Interspeech 2019



Agenda



METRICS



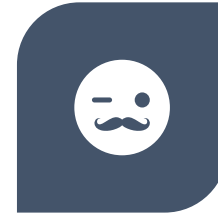
MEASUREMENTS



SPELLER FOR NBEST
RANKING



GEE ENTITY RESOLVER
FOR NBEST RANKING



APSAT FOR NBEST
RANKING



RECOVERY THROUGH
DISPLAYING MULTIPLE
RESULTS



INCREASING NUMBER
OF CANDIDATES

Metrics

Metrics

- **WER** (Levenshtein)
- **SIM** (similarity)
- New: **CI** (catastrophic impact on user understanding).
 - It is likely to happen when WERpw (WER per word in a query) ≥ 0.5
- Accuracy – absolute match with transcribed query

Data

- 1000 samples, Cortana traffic
- 1250 samples, Bing traffic
- Transcribed by vendors, corrected by FTE

~1K	SIM	WER	CI (%)	Accuracy
	0.9776	1.1009	9.74	83%

Terms

- **SR** – Speech recognition
- **SR candidate (NBest)** – candidates provided by SR. Typically, **NBest0** is suggested as best
- **Speller suggestion** – first suggestion that Bing speller provides for an SR candidate. Speller also may have its own candidates for suggestion, with different level of confidence
- **Speller Ranker** – one based on patterns of Speller suggestions
- **NBest0 Ranker** – one based on NBest Order from SR
- **Vote** – scores assigned by Speller Ranker to an SR candidate
- **“Misspelled”** - not a traditional notion that the utterance is misspelled in the sense of traditional English Grammar. Rather, Speller API returns a sound alternative (so called “suggestion”).

The story of Flipper (in PROD)

Trigger: Nbest-0 is misspelled, highest confidence suggestion Ω

There exists Nbest-k that contains Ω (PROD)

- *Variant*: There exists Nbest-k that is equal to Ω
- That generalizes to a Ranker:
 - Generate all Speller suggestions to each {Nbest-k | k in N}
 - [Reduce], e.g., by voting
 - Rank
 - Select the highest score Nbest-New that may not be in the list of SR candidates at all
 - Underlying ML problem is to minimize loss (CI (NBest-New))

Flipper Classifier Performance

- Based on the Bing Speller
- Decides whether SR NBest should be switched to another SR candidate
- Data below are for the 4-day run in PROD

BestCandidateIndex	Count	Description
0	87417	
-1	23883	Flipper didn't gain control (SR Error)
-2	19918	Flipper (Speller) timeout
1	1566	Flipped to a new index
2	537	
3	333	
4	170	
Performance (%)	2.89	Excluding SR errors and speller t/o

Improving SR
quality: Bing
Nbest
classifier

Total queries	133824
Modified queries total	2606
Estimated precision error rate of Bing Nbest classifier	3.15%
Estimated E2E impact, same answer, no speller suggestions	1.82%
Estimated E2E impact, better answer	0.33%

- Combining Bing Speller and SR NBest suggestions
- Perf cost: limited to 85ms
- Data below is for 4 days in PROD

Speller for NBest ranking

Call Speller for each NBest
candidate

Count number of votes for
each spelled candidate

Approach 1:

Choose the one with the
highest vote

Approach 2:

Choose candidates with >
1 vote

Let final ranking decision
to be done later

Speller and NBest0 Rankers agree

idx	SR candidate	nbestcorrect	rankercorrect	Speller suggestion	Transcribed query	Vote
0	where is the nearest public bathroom	1	1	where is the nearest public bathroom	where is the nearest public bathroom	0.5
1	wheres the nearest public bathroom	1	1	where's the nearest public bathroom	where is the nearest public bathroom	0
2	where is the nearest public bathrooms	1	1	where is the nearest public bathrooms	where is the nearest public bathroom	0
3	where is the nearest publicx bathroom	1	1	where is the nearest public bathroom	where is the nearest public bathroom	0.5
4	where is the nearest public bath room	1	1	where is the nearest public bathroom	where is the nearest public bathroom	0.5

Speller Ranker wins (based on CI by 1.94% based on accuracy 2.12%)

idx	SR candidate	nbestcorrect	rankercorrect	Speller suggestion	Transcribed query	Vote
0	closest papa johns pizza	0	1	closest papa john's pizza	closest papa john's pizza	0.5
1	closest papa john s pizza	0	1	closest papa john's pizza	closest papa john's pizza	0.5
2	closest papa johns pizzas	0	1	closest papa john's pizza	closest papa john's pizza	0.5
3	closest papa johns pete	0	1	closest papa john's pete	closest papa john's pizza	0
4	closest papa johns p	0	1	closest papa john's p	closest papa john's pizza	0

NBest0 Ranker wins

idx	SR candidate	nbestcorrect	rankercorrect	Speller suggestion	Transcribed query	Vote
0	find me a bang and olufsen tv	1	0	find me a bang and olufsen tv	find me a bang and olufsen tv	0
1	find me a bangin alofsin tv	1	0	find me a bangin alofsin tv	find me a bang and olufsen tv	0
2	find me a bangin olofsson tv	1	0	find me a bangin olofsson tv	find me a bang and olufsen tv	0.25
3	find me a bangin olofsson t v	1	0	find me a bangin olofsson tv	find me a bang and olufsen tv	0.25
4	find me a bangin olofsson teevee	1	0	find me a bangin olofsson teevee	find me a bang and olufsen tv	0

Recovery: Entity Resolver: Bang&Olufsen TV

Speller and NBest0 are both wrong

idx	SR candidate	nbestcorrect	rankercorrect	Speller suggestion	Transcribed query	Vote
0	what is the vesper wrestling club	0	0	what is the vesper wrestling club	what is the vesper rolling club	0
1	what is the vesper rolling club	0	0	what is the vesper rolling club	what is the vesper rolling club	0
2	what is the vesper rowing club	0	0	what is the vesper rowing club	what is the vesper rolling club	0
3	what does the vesper rowing club	0	0	what does the vesper rowing club	what is the vesper rolling club	0
4	what is the vesper wrestling club	0	0	what is the vesper wrestling club	what is the vesper rolling club	0

Recovery: Entity Resolver: Vesper Rolling Club not found

Speller and NBest0 are both wrong, revisited

idx	SR candidate	nbestcorrect	rankercorrect	Speller suggestion	Transcribed query	Vote
0	what is the vesper wrestling club	0	0	what is the vesper wrestling club	what is the vesper rolling club	0
1	what is the vesper rolling club	0	0	what is the vesper rolling club	what is the vesper rolling club	0
2	what is the vesper rowing club	0	0	what is the vesper rowing club	what is the vesper rolling club	0
3	what does the vesper rowing club	0	0	what does the vesper rowing club	what is the vesper rolling club	0
4	what is the vesper wrestling club	0	0	what is the vesper wrestling club	what is the vesper rolling club	0

Recovery: Entity Resolver: Vesper Rolling Club... also likely a transcriber's error: **Vesper Rowing Club**

Reason: There is a candidate that is identical to the second Speller suggestion, with a competitive regression score

GEE entity
resolver for
NBest ranking

Look up entities for
(remaining) NBest
candidates

Choose the longest
candidates(s) with
most words as entities

Generic Entity Extraction for Queries: What is the goal?

- Identify all entities in a **query**. Return their **spans**, **repository IDs** and their **type**.

system of a down hypnotize

Some entity repositories

Wikipedia (4.5m articles)

Freebase (20m+ entities)

Satori (500m+ entities)

System of a Down

Rock Band



System of a Down, also known by the acronym SOAD and often shortened to System, is an American four-piece rock band from Southern California, formed in 1994. The band currently consists of four Armenian-Americans: Daron Malakian (vocals, guitar), Serj Tankian (lead vocals, occasional keyboards and rhythm guitar), Shavo Odadjian (bass, background vocals) and John Dolmayan (drums).
en.wikipedia.org

Member	Member	Performance Role	Start	End
	Daron Malakian	Vocals Guitar	1994	

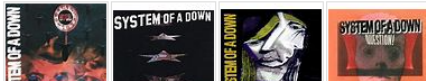
108cfe07-fb13-b951-0997-4badbe8161ef

Hypnotize

System of A down Song



"Hypnotize" is the lead single for Armenian American rock band System of a Down's album of the same name, which was released on November 22, 2005 (see 2005 in music). The video was filmed on September 28, 2005, at the Van Andel Arena in Grand Rapids, Michigan. It reached number one on Billboard's Hot Modern Rock Tracks chart and is the band's biggest international hit.
en.wikipedia.org

Composer	Daron Malakian
Track	Hypnotize
Related	

cdd971f7-a523-2890-3c07-2aaa7d98edeb

GEE Entity Resolver example

idx	SR candidate	nbestcorrect	rankercorrect	Speller suggestion	Transcribed query	Vote
	who scored the most points in the toronto raptors game last 0 night	1		who scored the most points in the 0 toronto raptors game last night	who scored the most points in the toronto raptors game last night	0
	whos scored the most points in the toronto raptors game last 1 night	1		whos scored the most points in the 0 toronto raptors game last night	who scored the most points in the toronto raptors game last night	0
	who is scored the most points in the toronto raptors game 2 last night	1		who is scored the most points in the 0 toronto raptors game last night	who scored the most points in the toronto raptors game last night	0
	who scored the most points in the durango raptors game last 3 night	1		who scored the most points in the 0 durango raptors game last night	who scored the most points in the toronto raptors game last night	0.25
	who scored the most points in the derango raptors game last 4 night	1		who scored the most points in the 0 durango raptors game last night	who scored the most points in the toronto raptors game last night	0.25

Decider: entity “Toronto Raptors”, rather than “Durango Raptors”

APSAT for NBest ranking

Look up predicted APSAT for
(remaining) NBest
candidates

Choose the one with the
highest value (if the
difference is statistically
significant)

What is APSAT?

Answer Prominence SATisfaction

Purpose Measures answer satisfaction at Pole/Top-1/GoBig based on user engagement with the whole page

Intuition A prominent answer must satisfy the majority (50%+) of users. If more than 50% of users end up satisfied by another answer/algo, or abandoned without satisfaction, then it's a DSAT for the prominent answer

Assumptions We assume that when there aren't any page clicks, the user has seen the prominent answer given its placement on the page, so abandonment can apply

Usage Rankers predict APSAT for every answer to determine answer placement on the page

Answers ship criteria Existing answers - APSAT > 0. New Answers – APSAT WinRate > 50%

Note: APSAT only applies to answers at TOP, Surplus applies to everything else

TOP-1
APSAT

TOP-2
Surplus

MOP
Surplus

BOP
Surplus

The screenshot shows a Google search for 'justin bieber'. The results are categorized into News, Videos, and Maps. The News section includes articles from 'bing.com/news' about Justin Bieber's relationship with Rita Ora, a Super Bowl commercial, and a song about Selena Gomez. The Videos section shows several clips from Justin Bieber's channel. Below the videos are links to his official website, IMDb profile, and social media presence on Twitter and Facebook. At the bottom, there are links to his Wikipedia page, Biography.com profile, Rotten Tomatoes page, and a section for pictures, videos, and breaking news. Related searches for Justin Bieber are listed at the very bottom.

APSAT Ranker

- Example: query is not resolved by Speller/Entity Resolver
 - “eye drop Madison”

Estimated combined CI impact improvement due to Entity Resolver and APSAT is 1.2%.

	Predicted APSAT
Final Prediction	52.55% <div></div>
	Predicted APSAT
Final Prediction	36.08% <div></div>

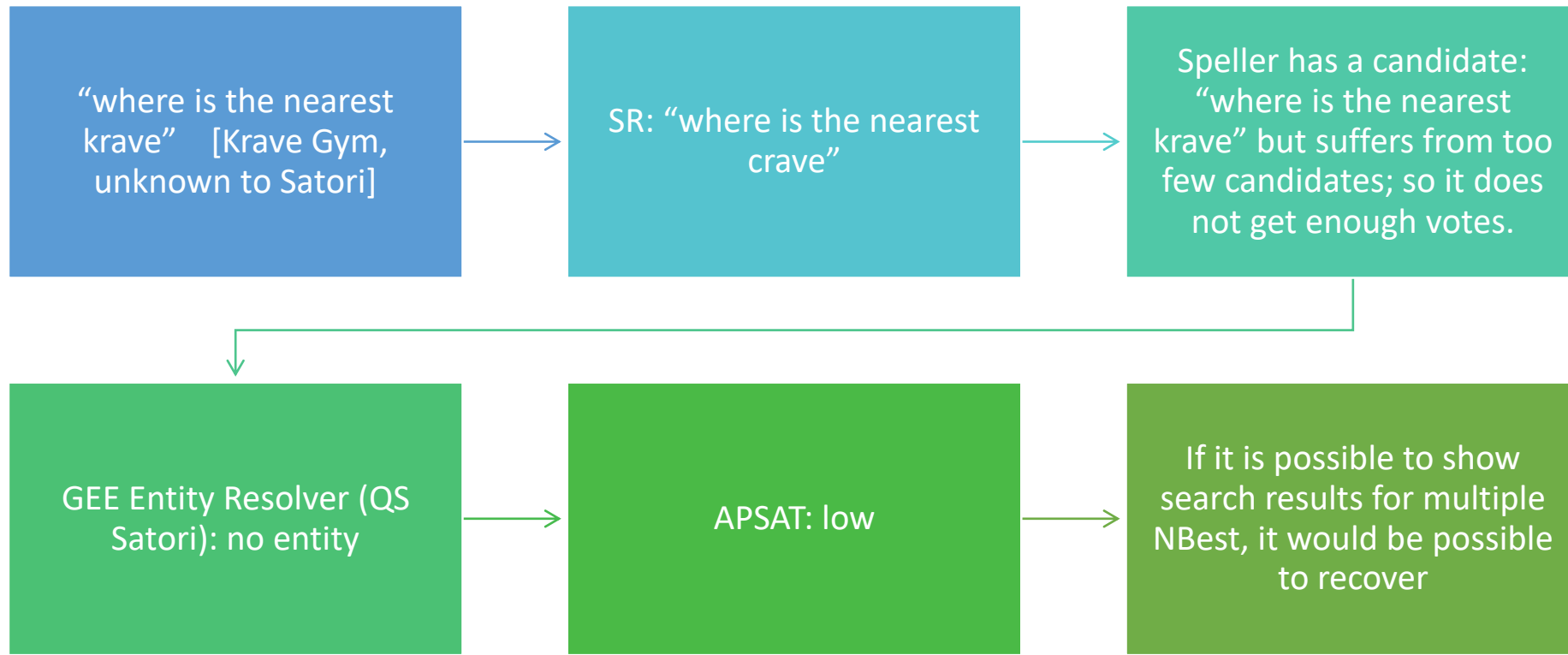
▼ QueryParses

Entities				Metadata	Confidence
CommonNounEntity	eye drop	0.938329	a1cd632c-170f-c4c4-2d7a-eb2429c2f03b		1
CommonNounEntity	medicine	0.796023	92ed15ca-802a-af7a-fabe-363fa96e96b1		

▼ QueryParses

Entities				Metadata	Confidence
CommonNounEntity	eye drop	0.926315	a1cd632c-170f-c4c4-2d7a-eb2429c2f03b		241
SatoriEntity	madison	0.590615	65426cc9-93eb-4569-8c2e-3b7c	location.city organization	

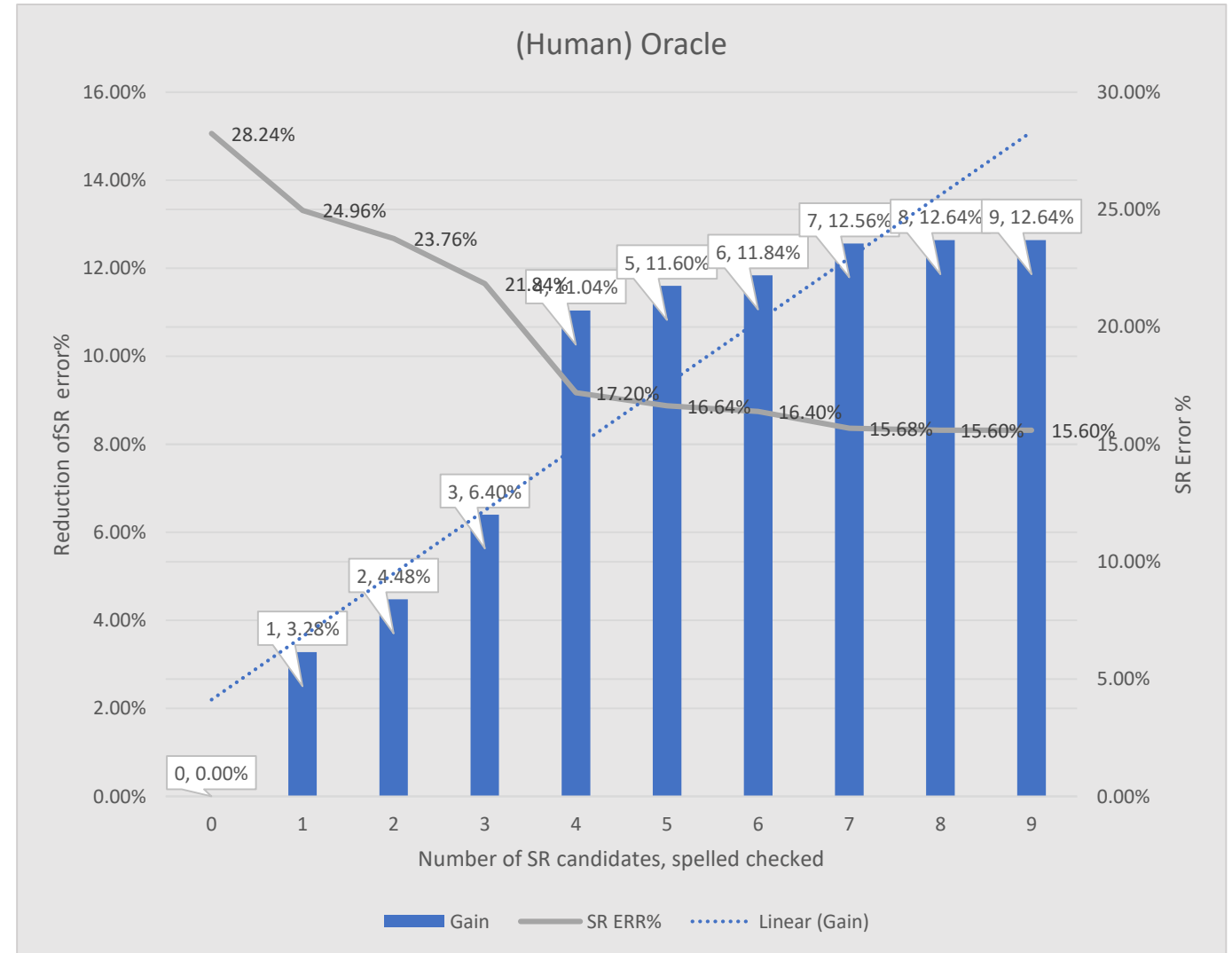
Gaps



More
candidates
improve
recovery

- **<N> candidates for NBest**
 - Currently fixed to N=5
 - Potential to get ~25 per query from SR (up to 50)
 - Early experimentation results: N>5, in general, and should be optimized per record (e.g., Classifier N=foo(words in query), confidence, ranker scores)
- **<M> suggestions for Speller**
 - Up to 50
 - Use regression and lambda scores
- **(Additional) Homophones**

(Human) Oracle



Improving SR quality

- SR team stopped using speller to meet compliance
- Bing Speller suggestions in Voice search results look confusing

fidel go campground anna cortis Washington

