

# Open Source Question Answer System

Advisor

Dr. Chris Pollett

Committee

Dr. Mark Stamp

Dr. Robert Chun

By

Salil Shenoy



# Outline

- Problem Statement
- Background on Question Answering Systems
- Implementation
- Experiments
- Conclusion

## Problem Statement

- Improve an existing English Question Answering System in Yioop
- Implement a Hindi Question Answering System for Yioop

# English Question Answer System

Calculated in 0.22066 seconds. Showing res

[Justin Trudeau is Condescending & Smug Just Like Obama - YouTube](#)  
www.youtube.com/watch?v=2oJ8\_oh7q1M Possible Answer: condescending smug Words: trudeau  
sign loading queue video  
Justin Trudeau is Condescending Smug Just Like Obama - YouTube. Sign in. Loading. ... Close. Undo  
Close. Watch Queue Queue. Watch Queue Queue. The next video is starting stop. Loading.  
[Cached.](#)

[Harper blasts Trudeau's economic plan as all 'unicorns and rainbows' - The Globe and Mail](#)  
www.theglobeandmail.com/news/pol...26862728 Possible Answer: truth Words: globe life riding  
harper business  
Harper blasts Trudeau's economic plan as all 'unicorns and ... Canada's Prime Minister and Conservative  
leader Stephen ... Canada's Prime Minister and Conservative leader Stephen ... Stephen Harper is calling  
Justin Trudeau's economic plan

[Trudeau and Cameron meet at G20 summit | Watch News Videos Online](#)  
globalnews.ca/video/2340452/trud...0-summit Possible Answer: the prime minister Words: trudeau  
justin national health minister  
Trudeau and Cameron meet at G20 summit. Watch ... Contests Live. Trudeau and Cameron meet at G20  
summit. Prime ... Minister Justin Trudeau and U.K. Prime Minister David Cameron  
[Cached.](#)

[Meet Justin Trudeau's new Liberal cabinet - Politics - CBC News](#)  
www.cbc.ca/news/politics/trudea-...3304176 Possible Answer: canada s 23rd Words: reuters wattie  
chris minister ontario  
Meet Justin Trudeau's new Liberal cabinet - Politics - CBC ... CBC News Politics. Trudeau's first day as  
prime minister 4:34. Justin Trudeau signals new style on 1st day as

[Justin Trudeau has a message to young voters - "voting matters" | Watch News Videos Online](#)  
globalnews.ca/video/2172012/just...matters Possible Answer: the prime minister Words: canada  
trudeau global national justin  
Justin Trudeau has a message to young voters voting ... at St. Canada. Justin Trudeau has a message to  
young voters ... Liberal leader Justin Trudeau sat down with ET Canada s  
[Cached](#)

# Hindi Search in Yioop...

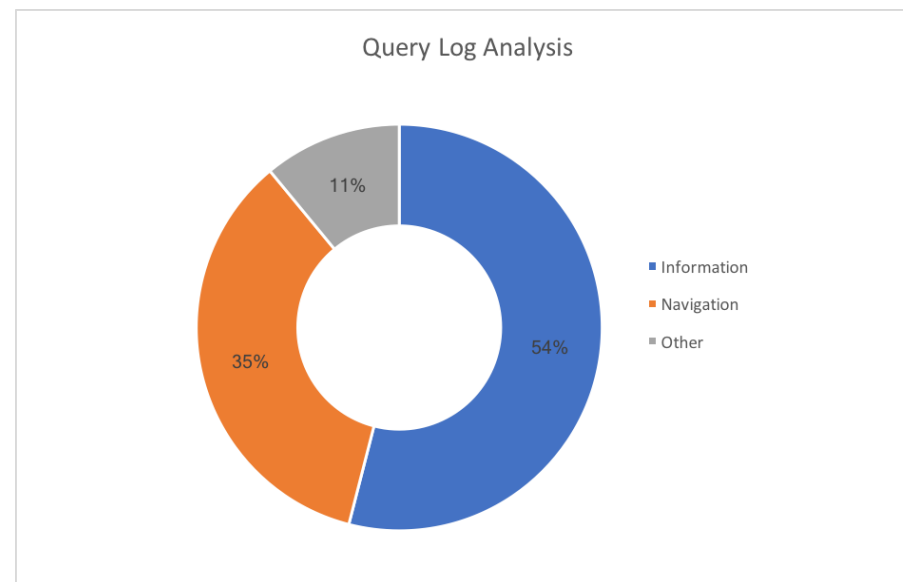
Web Images Videos News

 महात्मा गांधी कौन थे 

1.65137 seconds. Showing 1 - 10 of 196

[विहिप नेता का महात्मा गांधी को लेकर वि](#)  
khabar.ndtv.com/news/india/vhp-l...i-388146 Words: images ndtv advertisement vhp pm  
विहिप नेता का महात्मा गांधी को लेकर विवादस्पद बयान, कहा, डेढ़ पसलीवाला राष्ट्रपिता नहीं हो सकता - NDTV. NDTV. Business  
Hindi Movies. Cricket. Good Ti  
Cached. Similar. Inlinks. IP:23.212.52.59. Score:9.78  
[महात्मा गांधी को लेकर विवादित बयान - त?](#)  
khabar.ndtv.com/topic/महात्मा-गा...दित-बयान Words: ndtv images advertisement bipasha hindi  
महात्मा गांधी को लेकर विवादित बयान - ताज़ा खबर, वीडियो, महात्मा गांधी को लेकर विवादित बयान Hindi ब्रेकिंग न्यूज़, NDTV  
India. NDTVKhbar.com - NDTV.  
Cached. Similar. Inlinks. IP:23.212.52.59. Score:9.67  
[महात्मा गांधी - ताज़ा खबर, वीडियो, महा?](#)  
khabar.ndtv.com/topic/महात्मा-गा...मा-गांधी Words: pm gandhi modi ist read  
Read More: मणिशंकर अय्यर , महात्मा गांधी , गांधी जयंती , नरेंद्र मोदी , सफाई अभियान , स्वच्छ भारत अभियान , Mani Shankar  
Aiyar , Mahatma Gandhi , Swach  
Cached. Similar. Inlinks. IP:96.17.109.16. Score:9.59  
[महात्मा गांधी जयंती - ताज़ा खबर, वीडि?](#)  
khabar.ndtv.com/topic/महात्मा-गा...धी-जयंती Words: gandhi pm read ist india  
Read More: मणिशंकर अय्यर , महात्मा गांधी , गांधी जयंती , नरेंद्र मोदी , सफाई अभियान , स्वच्छ भारत अभियान , Mani Shankar  
Aiyar , Mahatma Gandhi , Swach  
Cached. Similar. Inlinks. IP:23.212.52.59. Score:9.48  
[Mahatma gandhi - ताज़ा खबर, वीडियो, Mahatma gandhi Hindi ब्रे](#)  
khabar.ndtv.com/topic/mahatma-gandhi Words: gandhi pm mahatma ist modi  
Mahatma gandhi - ताज़ा खबर, वीडियो, Mahatma gandhi Hindi ब्रेकिंग न्यूज़, NDTV India. Mahatma Gandhi.  
Mahatma gandhi. Read More: महात्मा गांधी को श्र  
Cached. Similar. Inlinks. IP:23.212.52.59. Score:9.32  
[रामदेव की भूमिका महात्मा गांधी और जेप?](#)

# What type of questions are asked...



# Existing Question Answer Systems...



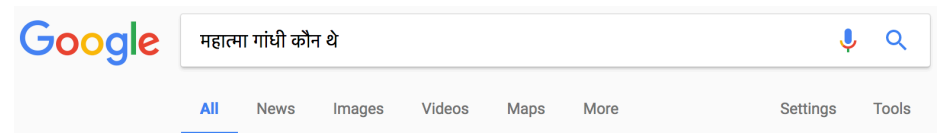
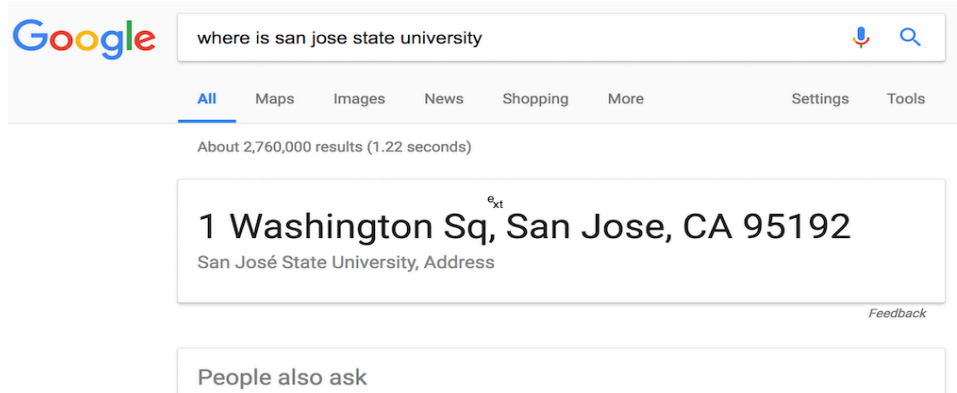
==> Who directed Gone with the Wind

**Gone with the Wind (1939)** was directed by Victor Fleming, George Cukor, and Sam Wood.

I know about one more movie called "Gone With The Wind": **Gone with the Wind (1954)**

Source: [The Internet Movie Database](#)

[Back to home page](#)



About 4,530,000 results (0.55 seconds)

**महात्मा गांधी - विकिपीडिया**

[https://hi.wikipedia.org/wiki/महात्मा\\_गांधी](https://hi.wikipedia.org/wiki/महात्मा_गांधी) [Translate this page](#)

मोहनदास करमचंद गांधी (2 अक्टूबर 1869 - 30 जनवरी 1948) भारत एवं भारतीय स्वतंत्रता आंदोलन के एक प्रमुख राजनैतिक एवं आध्यात्मिक नेता थे। वे सत्याग्रह (व्यापक सविनय अवज्ञा) के माध्यम से अत्याचार के प्रतिकार के अग्रणी नेता थे, उनकी इस अवधारणा की नींव सम्पूर्ण अहिंसा के सिद्धान्त पर रखी गयी थी जिसने भारत को आजादी दिलाकर पूरी दुनिया में जनता के नागरिक अधिकारों एवं स्वतंत्रता के प्रति आन्दोलन के लिये प्रेरित किया। उन्हें दुनिया में आम जनता **महात्मा गांधी** के नाम से जानती ...

[महात्मा गांधी की हत्या](#) · [महात्मा गांधी का ...](#) · [देवदास गांधी](#)

**Poem on Mahatma Gandhi in Hindi | हिन्दी कविता - YouTube**

<https://www.youtube.com/watch?v=YpXdE2wfJlg>

Sep 29, 2016 - Uploaded by बोलते चित्र

Poem on **Mahatma Gandhi** in Hindi | हिन्दी कविता- राष्ट्रपिता महात्मा गांधी पे पर गांधी जयंती पे कविता This is short and simple poem for ...

**महात्मा गांधी की हत्या - विकिपीडिया**

[https://hi.wikipedia.org/wiki/महात्मा\\_गांधी\\_की\\_हत्या](https://hi.wikipedia.org/wiki/महात्मा_गांधी_की_हत्या) [Translate this page](#)

मोहनदास करमचंद गांधी की हत्या 30 जनवरी 1948 की शाम को नई दिल्ली स्थित बिड़ला भवन में गोली मारकर की गयी थी। वे रोज शाम को प्रार्थना किया करते थे। 30 जनवरी 1948 की शाम को जब वे संध्याकालीन प्रार्थना के लिए जा रहे थे तभी नाथूराम गोडसे नाम के व्यक्ति ने पहले उनके पैर छुए और फिर सामने से उन पर बैरेट पिस्तौल से तीन गोलियाँ दाग दीं। उस समय गान्धी अपने अनुचरों से घिरे हुए थे। इस मुकदमे में नाथूराम गोडसे सहित आठ लोगों को हत्या की साजिश में आरोपी बनाया गया था। इन आठ लोगों में ...

[हत्या के दिन](#) · [सुरक्षा](#) · [अधूरी रह गयी ...](#) · [पत्र](#)

# Question Answer Systems

- Types of Question Answering Systems
  - Closed Domain  
eg: BASEBALL, LUNAR, IRSL etc.
  - Open Domain  
eg: START, Google, etc.



# Question Answer Systems Paradigms

- Text Based / IR Based
- Knowledge Based
- Hybrid

## Part of Speech Tagging

- Approaches to PoS tagging,
  - Rule Based (Brill Tagger)
  - Machine Learning
- For our system, implemented variant of Brill tagger for Hindi PoS tagging

## Part of Speech Tagging

- Input is sentence processed, stemmed, n-gram, stop word and punctuation removed
- Terms present in the lexicon in database are tagged and words not found are tagged as unknown
- PoS tagging rules are applied for unknown words

## PoS tagging Example

- Example :

- महामा गाँधी का जम 2 अक्टूबर को हुआ

- महामा ~ NN गाँधी ~ NN का ~ IN जम ~ NN 2 ~ QT  
अक्टूबर ~ NN हुआ ~ VB

## Parse Tree Generation

- For Hindi, we use following grammar rules to extract triplets

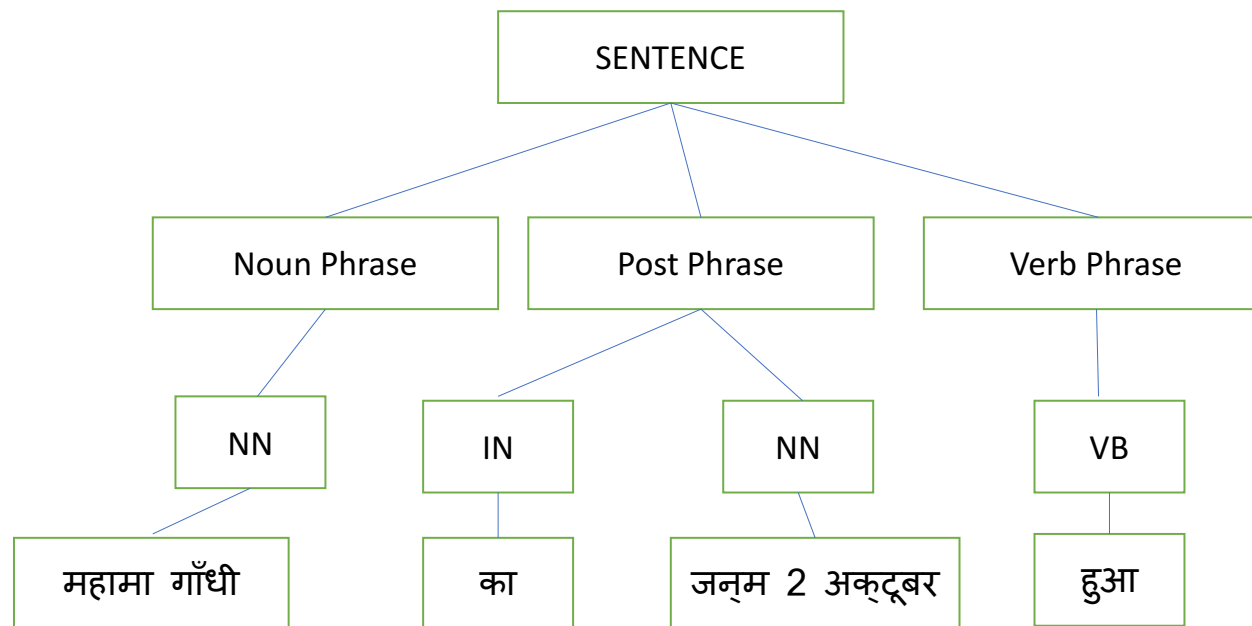
**S:** {< **NP** >< **PP** >< **VP** >}

**NP:** {< **JJ** | **NN\*** >}+

**PP:** {< **IN** | **JJ** | **NN** | **PP** >}+

**VP:** {< **VB\*** >< **IN** | **NP** >}

# Parse Tree



# Triplet Extraction

- Input is the parse tree
- Two types of triplets
  - Concise
  - Raw
- For Hindi, triplet is -> [Subject – Object - Verb]

# Triplet Extraction

- For parse tree we saw on the previous slide the triplets and the corresponding answers are as follows,
  - [महामा गाँधी - qqq - हुआ ] => जन्म 2 अक्टूबर
  - [qqq - जन्म 2 अक्टूबर - हुआ ] => महामा गाँधी
  - [महामा गाँधी - जन्म 2 अक्टूबर - qqq] => हुआ



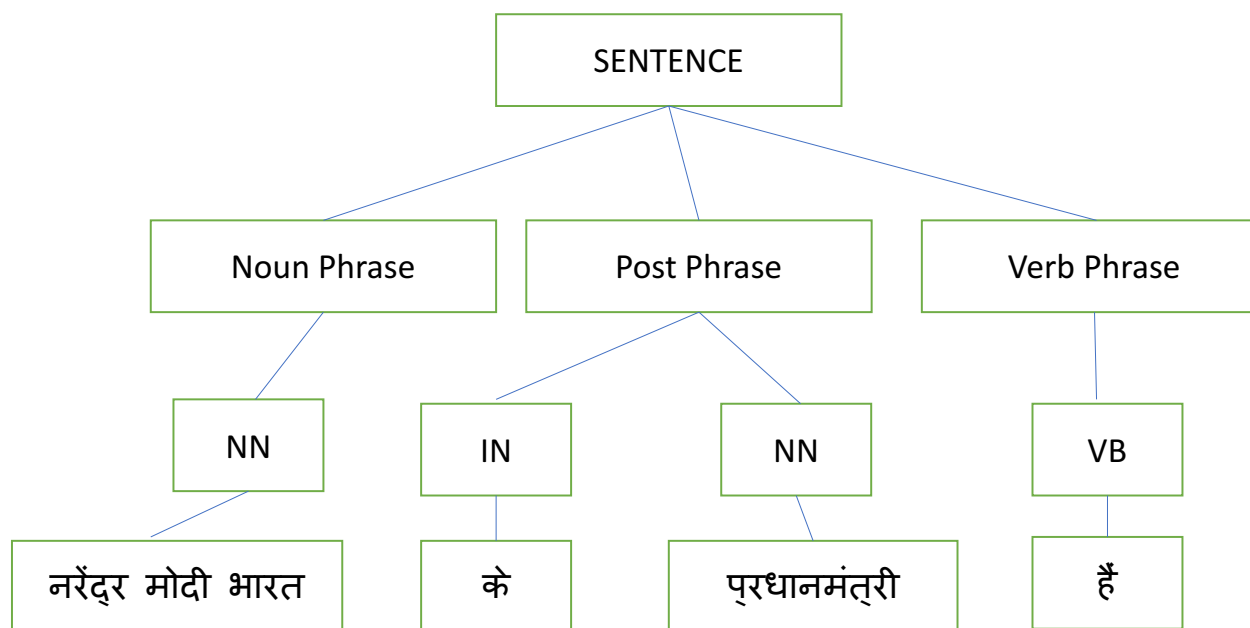
## PoS tagging Example

- Example :

- नरेंद्र मोदी भारत के कतिने प्रधानमंत्री हैं

- नरेंद्र ~ NN मोदी ~ NN भारत~NNP के ~ IN  
प्रधानमंत्री ~ NN हैं ~ VB

# Parse Tree



- For parse tree we saw on the previous slide the triplets and the corresponding answers are as follows,

- [नरेंद्र मोदी भारत - **qqq** - हैं] => प्रधानमंत्री
- [**qqq** - प्रधानमंत्री - हैं] => नरेंद्र मोदी भारत
- [नरेंद्र मोदी भारत - प्रधानमंत्री - **qqq**] => हैं

# Named Entities

- Named entity recognition, one of the major component of a QA system
- Ways to recognizing entities in text...
  - Maintain a handwritten dictionary of entities
  - Use automated methods for eg. Using Wikipedia to recognize entities.

# Named Entities


- User can add, delete, view entities for a selected locale

Crawl Time	Search Time	Named Entity	Test Options
<h3>Add Named Entity</h3> <p><input checked="" type="radio"/> Add Entity <input type="radio"/> Upload File</p> <p>Entity <input type="text"/></p> <p>Locale: <input type="text" value="ar"/></p> <p><input type="button" value="Save"/></p>			

Crawl Time	Search Time	Named Entity	Test Options
<h3>Add Named Entity</h3> <p><input type="radio"/> Add Entity <input checked="" type="radio"/> Upload File</p> <p>File <input type="button" value="Browse..."/> No file selected.</p> <p>Locale: <input type="text" value="ar"/></p> <p><input type="button" value="Save"/></p>			

- User can view entities for a locale

[\[root\]](#) [Settings](#) [Sign Out](#)

 - Admin [Page Options]

Account Access

Manage Account

Manage Users

Manage Roles

Crawls

Manage Crawls

Manage Classifiers

Page Options

Results Editor

Search Sources

Web Scrapers

Social

Edit Entities : English

Entity Name	Action
Salil Shenoy	<a href="#">Delete</a>
College of Science	<a href="#">Delete</a>
Pune	<a href="#">Delete</a>
New Jersey	<a href="#">Delete</a>
Manchester	<a href="#">Delete</a>
London	<a href="#">Delete</a>
Family	<a href="#">Delete</a>

[Back](#)

# Experiments

- We created an index for English and another for Hindi by setting up a crawl on wikipedia pages.
- To setup a crawl in Yioop, under the Manage Crawls Tab, click on Options, add the seed sites
- For hindi, I used <https://hi.wikipedia.org/> and domains to crawl as 'co.in' and 'in'
- For english <https://en.wikipedia.org/> and domains to crawl as 'com'

# Experiments

Yioop without Hindi Q/A

Web Images Videos News

Yioop! महात्मा गांधी कौन थे

1.65137 seconds. Showing 1 - 10 of 196

[विहिप नेता का महात्मा गांधी को लेकर वि](#)  
khabar.ndtv.com/news/india/vhp-...i-388146 Words: [images](#) [ndtv](#) [advertisement](#) [vhp](#) [pm](#)  
विहिप नेता का महात्मा गांधी को लेकर विवादास्पद बयान, कहा, डेढ़ पसलीवाला राष्ट्रपिता नहीं हो सकता - NDTV. NDTV. Business Hindi Movies. Cricket. Good Ti  
[Cached](#) [Similar](#) [Inlinks](#) IP:23.212.52.59 Score:9.78  
[महात्मा गांधी को लेकर विवादित बयान - त?](#)  
khabar.ndtv.com/topic/महात्मा-गा...दित-बयान Words: [ndtv](#) [images](#) [advertisement](#) [bipasha](#) [hindi](#)  
महात्मा गांधी को लेकर विवादित बयान - ताज़ा खबर, वीडियो, महात्मा गांधी को लेकर विवादित बयान Hindi ब्रेकिंग न्यूज़, NDTV India. NDTVKhabar.com - NDTV.  
[Cached](#) [Similar](#) [Inlinks](#) IP:23.212.52.59 Score:9.67  
[महात्मा गांधी - ताज़ा खबर, वीडियो, महा?](#)  
khabar.ndtv.com/topic/महात्मा-गा...मा-गांधी Words: [pm](#) [gandhi](#) [modi](#) [ist](#) [read](#)  
Read More: मणिशंकर अय्यर , महात्मा गांधी , गांधी जयंती , नरेंद्र मोदी , सफाई अभियान , स्वच्छ भारत अभियान , Mani Shankar Aiyar , Mahatma Gandhi , Swach  
[Cached](#) [Similar](#) [Inlinks](#) IP:96.17.109.16 Score:9.59  
[महात्मा गांधी जयंती - ताज़ा खबर, वीडि?](#)  
khabar.ndtv.com/topic/महात्मा-गा...धी-जयंती Words: [gandhi](#) [pm](#) [read](#) [ist](#) [india](#)  
Read More: मणिशंकर अय्यर , महात्मा गांधी , गांधी जयंती , नरेंद्र मोदी , सफाई अभियान , स्वच्छ भारत अभियान , Mani Shankar Aiyar , Mahatma Gandhi , Swach  
[Cached](#) [Similar](#) [Inlinks](#) IP:23.212.52.59 Score:9.48  
[Mahatma gandhi - ताज़ा खबर, वीडियो, Mahatma gandhi Hindi ब्रे](#)  
khabar.ndtv.com/topic/mahatma-gandhi Words: [gandhi](#) [pm](#) [mahatma](#) [ist](#) [modi](#)  
Mahatma gandhi - ताज़ा खबर, वीडियो, Mahatma gandhi Hindi ब्रेकिंग न्यूज़, NDTV India. Mahatma Gandhi.  
Mahatma gandhi. Read More: महात्मा गांधी को श्र  
[Cached](#) [Similar](#) [Inlinks](#) IP:23.212.52.59 Score:9.32  
[रामदेव की भूमिका महात्मा गांधी और जेप?](#)

Yioop With Hindi Q/A

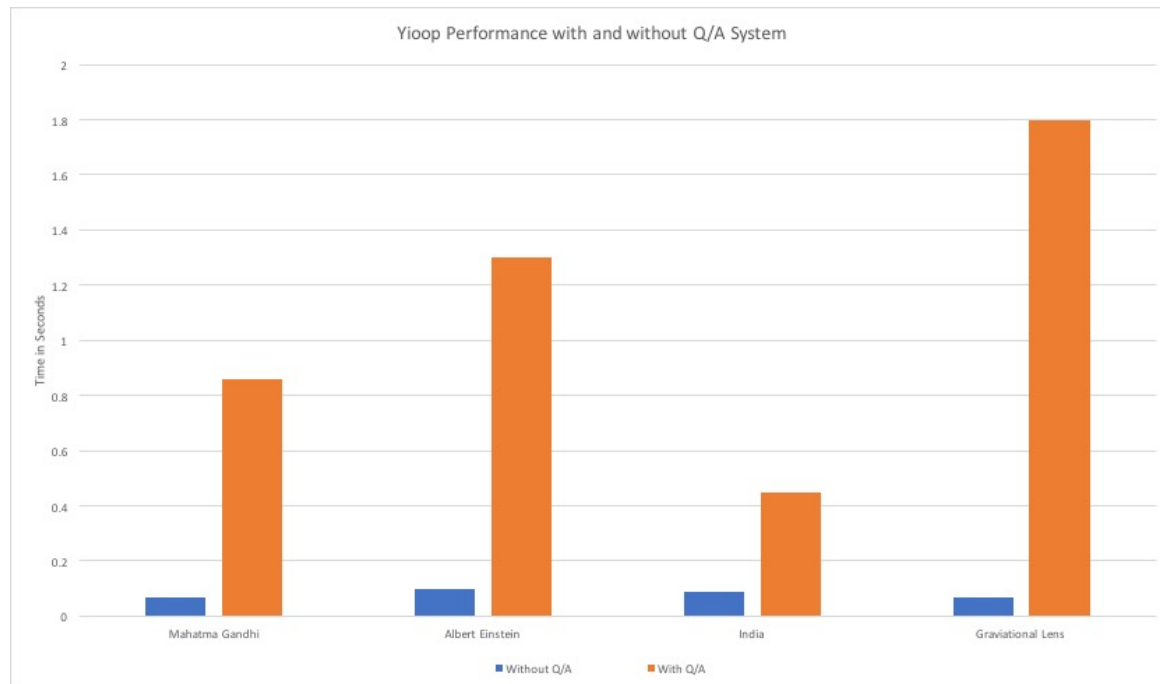
Yioop! महात्मा गांधी कौन थे

0.86897 seconds. Showing 1 - 10 of 292

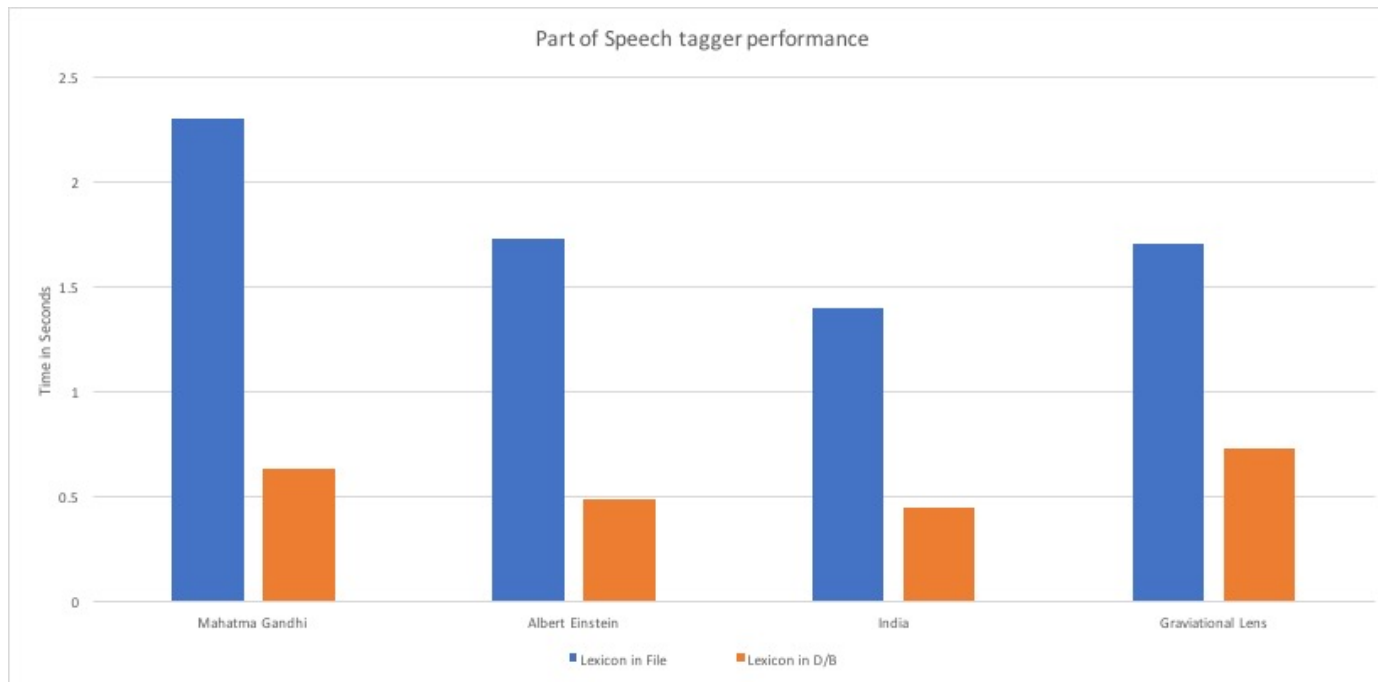
Search: [महात्मा गांधी कौन थे](#)  
संभावित जवाब महान स्वतंत्रता सेनानी  
[महात्मा गांधी - विकिपीडिया](#)  
hi.wikipedia.org/wiki/महात्मा\_गा...मा\_गांधी संभावित जवाब महान स्वतंत्रता सेनानी Words: [न](#) [व](#) [कर](#) [उन](#) [और](#)  
[Cached](#) [Similar](#) [Inlinks](#) IP:198.35.26.96 Score:10.0  
[देवदास गांधी - विकिपीडिया](#)  
hi.wikipedia.org/wiki/देवदास\_...82थी संभावित जवाब महान स्वतंत्रता सेनानी Words: [न](#) [व](#) [कर](#) [उन](#) [और](#)  
[Cached](#) [Similar](#) [Inlinks](#) IP:198.35.26.96 Score:9.44  
[कस्तूरबा गांधी - विकिपीडिया](#)  
hi.wikipedia.org/wiki/कस्तूर...82थी संभावित जवाब स्वतंत्रता संग्राम Words:  
[Cached](#) [Similar](#) [Inlinks](#) IP:198.35.26.96 Score:9.29  
[गांधी-इरविन समझौता - विकिपीडिया](#)  
hi.wikipedia.org/wiki/गांधी-इ...8Cता संभावित जवाब स्वतंत्रता संग्राम Words:  
[Cached](#) [Similar](#) [Inlinks](#) IP:198.35.26.96 Score:9.29  
[मणिलाल गांधी - विकिपीडिया](#)  
hi.wikipedia.org/wiki/मणिलाल\_...82थी संभावित जवाब उच्च शिक्षा Words:  
[Cached](#) [Similar](#) [Inlinks](#) IP:198.35.26.96 Score:9.07  
[साबरमती आश्रम - विकिपीडिया](#)  
hi.wikipedia.org/wiki/साबरमती...8Dरम संभावित जवाब महान स्वतंत्रता सेनानी Words: [न](#) [व](#) [कर](#) [उन](#) [और](#)  
[ीचरब नामक स्थान में महात्मा गांधी द्वारा हुई थी। सन् 1917 में यह आश ... थी। आश्रम का जीवन गांधी जी के सत्य, अहिंसा](#)



# Before / After Hindi Q/A integration

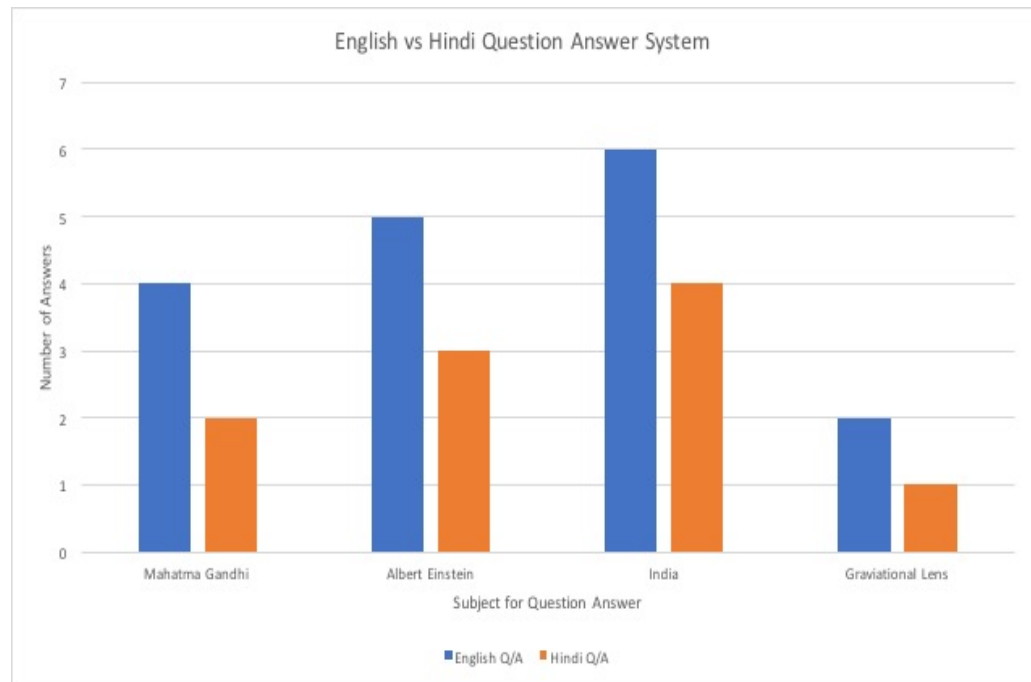


# Part of Speech tagger Performance

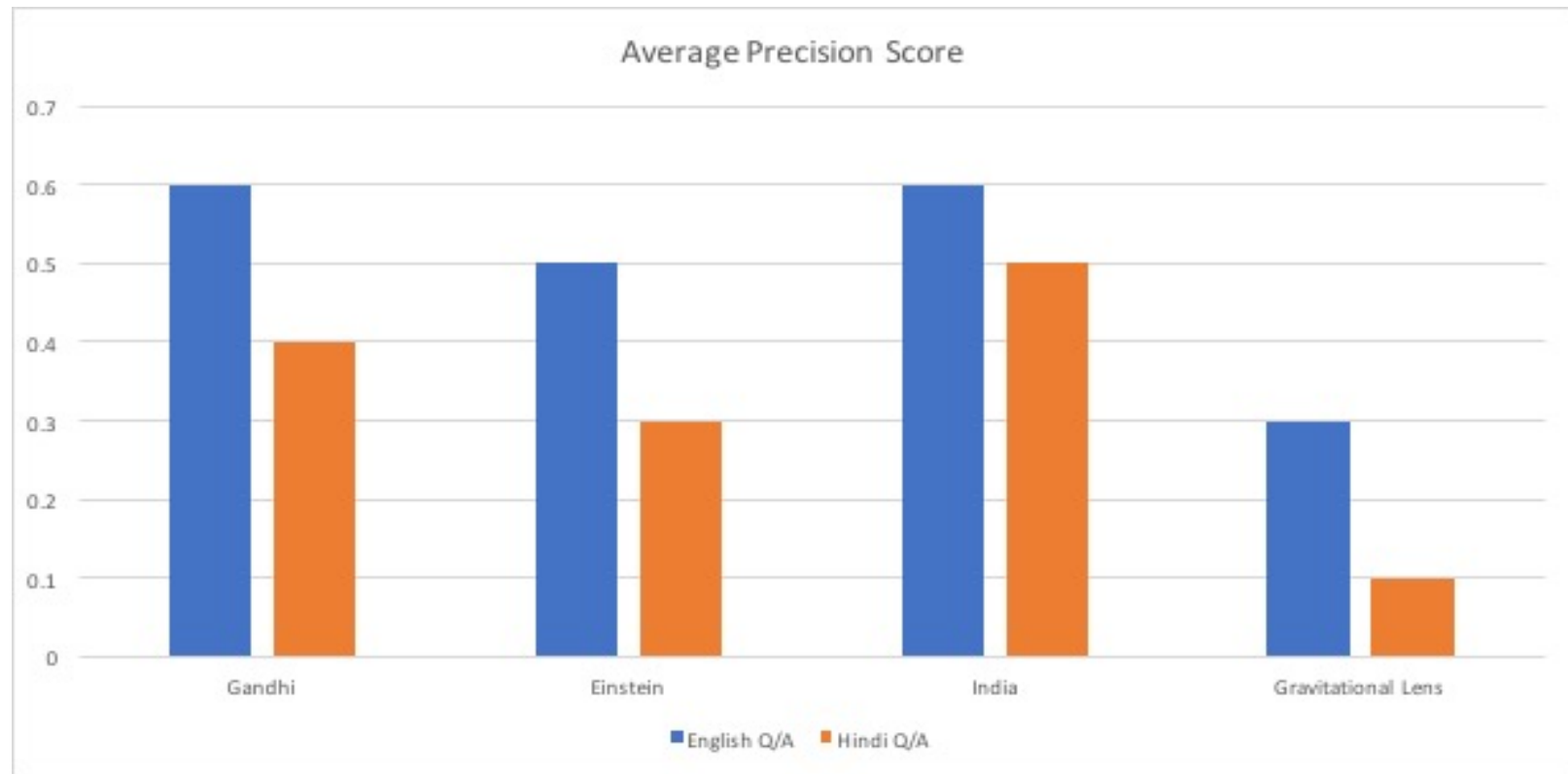


# English vs Hindi Q/A

- I used 4 topic to evaluate the retrieval efficiency of the system
- For a given topic, I asked the same set of questions to english and hindi systems.
- For eg:  
“Who was Mahatma Gandhi”  
“महात्मा गांधी कौन थे”  
“Who is Albert Einstein ?”  
“अल्बर्ट आइंस्टीन कौन थे”
- The accuracy is evaluated by comparing retrieved answer with known answer dataset



# Average Precision Score



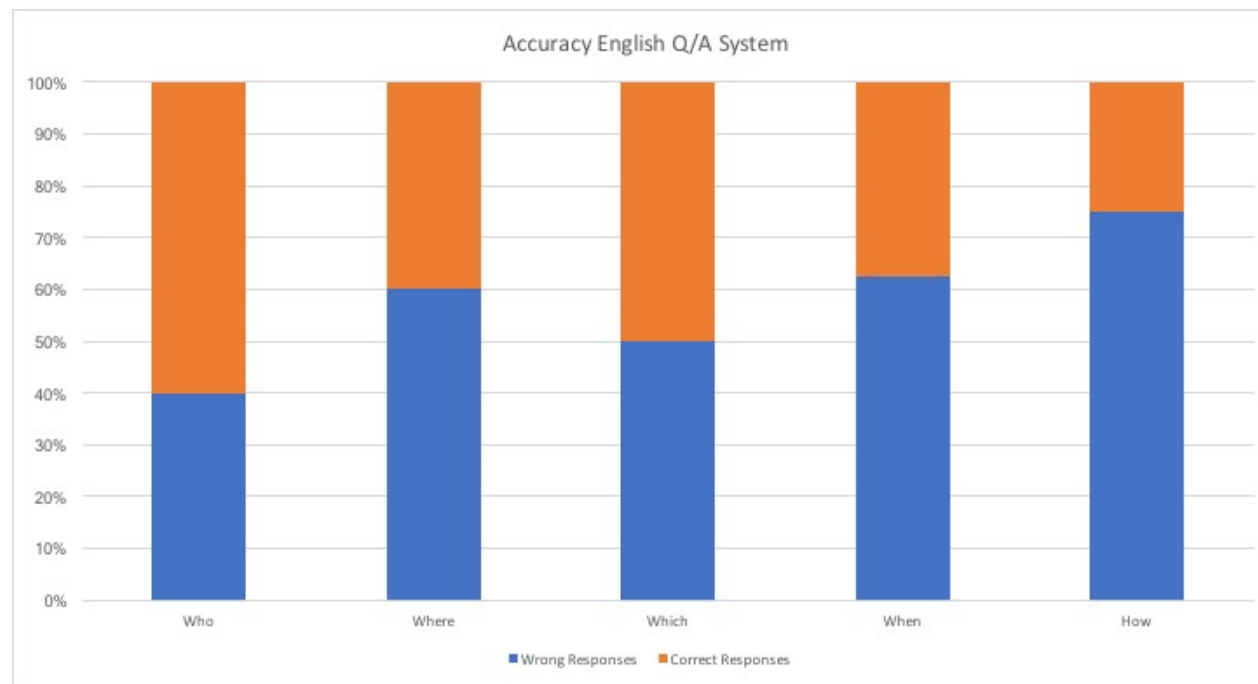
# Reciprocal Rank

Query ▼	Answers ▼	Correct Response ▼	Rank ▼	Reciprocal Rank ▼
महात्मा गांधी कौन थे	महान स्वतंत्रता सेनानी, उच्च शिक्षा, स्वतंत्रता संग्राम	महान स्वतंत्रता सेनानी	1	1
अल्बर्ट आइंस्टीन कौन है	अमेरिकी इंजीनियर, सफलता, डॉक्टर, सैद्धांतिक भौतिकविद्	सैद्धांतिक भौतिकविद्	4	1 / 4
गुरुत्वाकर्षण लेंस क्या है	भौतिकी, न्यूटन, किरणों, प्रभाव,	प्रभाव	15	1/15

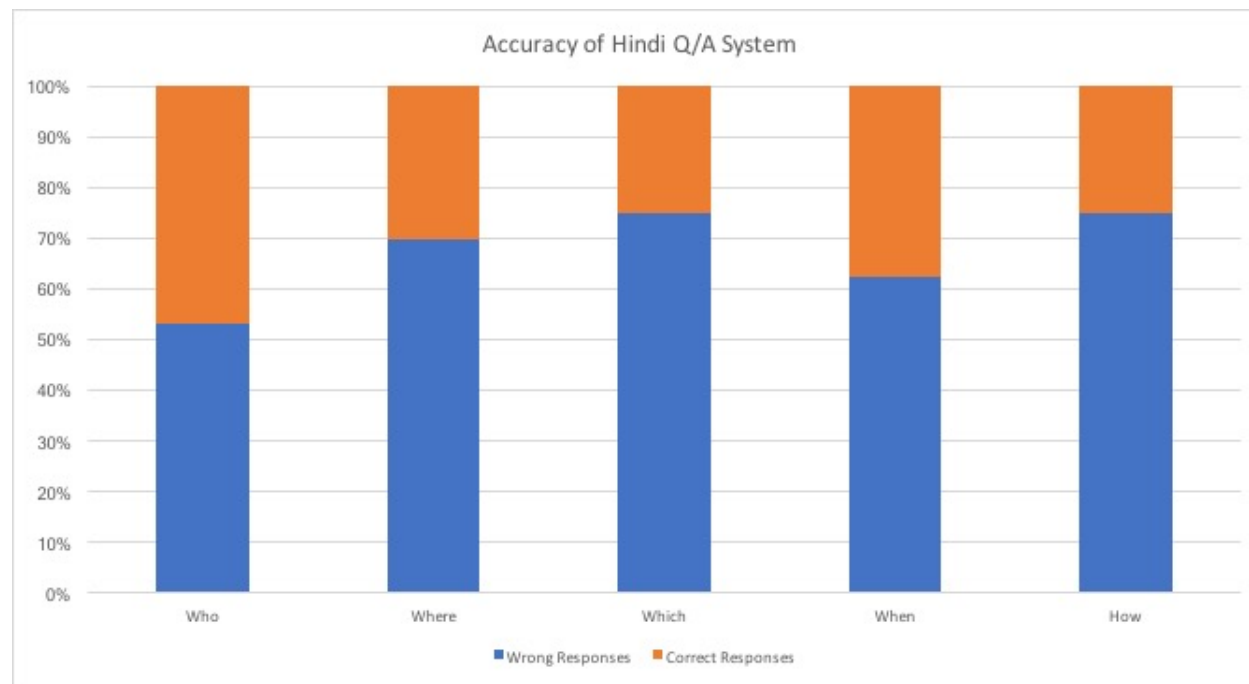
# Accuracy of Question Answering

- I created a set of 25 questions, corresponding set of answers known to be true
- I asked the questions from the question set and compared the retrieved answer with known answer set to evaluate efficiency

# Accuracy English Q/A



# Accuracy Hindi Q/A





Questions ?