Context Aware Web-mining Engine

Christopher Goh Zhen Fung
River Valley High School
Singapore
chrisgzf@gmail.com

Koay Tze Min
Raffles Institution
Singapore
tzemin.koay99@gmail.com

***Abstract*—**The context of a user's query is often lost when using search engines such as Google, as they limit the number of search terms that can be input. Therefore, they may not return the most relevant or desired results often. This project developed a context-aware web mining engine (CAWE), which allows the user to use the entire content of multiple documents as the query, to search and rank web documents by their similarity to it. CAWE combines the search results from Google, Yahoo and Bing search engines, and further crawls the links found in the downloaded web documents. The web documents are then ranked according to their content's relevance to the query documents. This ability for document matching is enabled by natural language processing techniques. Experiments showed that CAWE performed better than Google in finding more relevant web documents. CAWE has numerous plausible potential applications such as terrorist profiling for countering terrorism and disaster monitoring for humanitarian relief operations.

Keywords-natural language processing, machine learning, web-mining, data extraction, document search engine

I. Introduction

A. *Purpose of Research Area*

Since its invention, the Internet has become an unimaginably vast repository of information. By itself, this information is disorganized and messy, such that users have to rely on search engines such as Google to index and search through the web. While search engines have proved to be useful, they are not context-aware. For example, searching for a "crane" when you want to find out about the crane vehicle might net you results of the crane bird. Of course, the user can add keywords to give search engines more context, such as searching for "crane vehicle" instead. However, there is a limit to how much context it can be given (i.e. limited number of search terms allowed). The context of the query is thus often lost, and the returned results are not always the most relevant, and might miss the desired documents hidden within the links of the returned results. To find truly relevant information, a user often visits multiple websites and manually assesses each one - an extremely tedious and time-consuming process, especially if the content in these sites are lengthy.

This project aimed to develop an engine that lets users find documents in the Internet that are more relevant to their query, by ensuring that the context of their query is better captured. Titled the Context-Aware Web-mining Engine (CAWE), it extracts keywords from an arbitrary-length query from the user, before using the keywords on multiple search engines to search for the first layer of web documents. These web documents are then downloaded, with the URLs within them extracted to download more layers of web documents. These documents are ranked according to their similarity to the query, using natural language processing with machine learning techniques, to return search results of higher relevancy.

We compared the performance of CAWE with Google, the most widely used search engine in the world, holding 70.69% of the search engine market share as of November 2015 [1]. Most popular search engines (Bing, DuckDuckGo, Baidu, Yahoo, etc.) make use of similar PageRank techniques like Google, with variants of semantic matching [2].

B. *Drawbacks of Current Popular Search Engines*

- Popular search engines tend to rank web documents deemed most widely used first (social networking sites, organization websites, etc.). This can be useful in some cases, but not for research and information acquisition, where highly relevant information is sought after.
- Usually, only a limited-length query of up to 32 words can be input into the search engine, which makes document matching impossible.
- Many websites use search engine optimization techniques that can artificially inflate their search engine rankings

These properties give results that are not truly ranked based on relevance, and make it tedious for users to search through pages of results and identify relevant sites.

II. Research Materials and Methods

A. *Engine Creation*

CAWE was programmed in Java with IntelliJ IDEA, a Java IDE [3]. CAWE has two main phases, the Searcher and the Ranker as illustrated in Fig. 1.

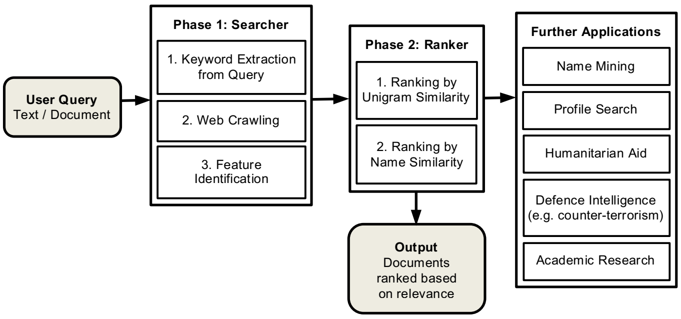


Figure 1. Flowchart of CAWE's Web-mining Process

1) Phase 1: Searcher

The searcher takes in an arbitrary-length query or query document(s), and outputs a list of web documents.

a) Keyword Extraction from Query

Keywords are needed to activate multiple search engines to provide our first layer of web documents. These keywords are extracted from the query documents. To do so, the contents of the query files are first tokenized into unigrams using the tokeniser from the Apache Lucene libraries. Stop words (insignificant words like 'the') are excluded. The unigrams are then ranked by importance, with the top 10 unigrams suggested to the user as keywords for search engines. The importance of each unigram is computed based on term frequency-inverse document frequency (TF-IDF) and a factor that indicates if it is a name entity.

TF-IDF is a measure of importance for a term in a collection of documents, where the higher the value of the term, the greater its importance. The TF-IDF of term *t* is calculated as shown in (1).

$$\mathrm{TFIDF}_t = \mathrm{TF}_t \times \mathrm{IDF}_t \quad (1)$$

Term Frequency (TF) refers to the number of times a term appears in a document. The greater the term frequency, the more important it is in that document. As different documents have different numbers of terms, the term frequencies are normalized as shown in (2).

$$\text{Normalised } \mathrm{TF}_t = \frac{\mathrm{TF}_t}{n} \quad (2)$$

where *t* is the term, *n* is the total number of terms in the document

Document frequency (DF) is the number of documents that a particular term appears in a collection of documents. Less important words such as "people" are so commonly used that they appear in almost every document, so their document frequencies are often higher than more important keywords occurring in fewer documents. Therefore, inverse document frequency (IDF) assigns higher weight to less frequently used terms. The IDF of a term *t* is calculated as shown in (3).

$$\mathrm{IDF}_t = \log\left(\frac{N}{\mathrm{DF}_t}\right) \quad (3)$$

where *N* is the total number of documents in the collection

In CAWE, the DF of words are pre-trained on 441803 text articles extracted from Wikipedia, the world's largest online encyclopaedia [4]. Also, the names of people, organizations and locations are given more weight than other terms. These names are extracted using the name entity recognizer (NER) from the Stanford CoreNLP libraries [5]. The user may enter these suggested keywords into CAWE, or they may come up with their own, to be used as the starting point for the web crawler.

b) Web Crawling

The input keywords are used to form search queries on 3 search engines (Google, Bing, Yahoo), where the searcher fetches URLs from every search result using the Jsoup HTML Parser [6]. Every URL fetched is added to the multithreaded processing queue, where raw HTML/PDF/Word Document content is downloaded from the URLs. All URLs from the raw documents are extracted and added to the processing queue again, to download their content in a breadth-first search manner. The depth of the crawling is configurable by the user.

c) Feature Identification

Content-tagged text content (e.g. paragraph-tagged HTML text content with <p>, header text, etc.) is extracted from the raw document with Jsoup (for HTML), Apache POI (for Microsoft Office) [7] and Apache PDFBox (for PDF) [8] for ranking. This removes most possibly irrelevant content such as navigation text or advertisements.

2) Phase 2: Ranker

The Ranker ranks the downloaded web documents according to each one's relevance to the query documents. The relevance is measured by calculating the cosine similarity between each document and the query. It takes into account two factors, unigram similarity and name similarity, before giving a final rank.

Cosine similarity is a measure of the similarity between two vectors of an inner product space that calculates the cosine of the angle between them [9]. In this case, the vectors represent the two documents to be compared. Each dimension of the vectors corresponds to an individual term in the documents, and the TF-IDF of each term is assigned as the values in the vectors, similar to the previous TF-IDF calculation method. The cosine similarity of two vectors, A and B, is given in (4).

$$\text{cosine similarity} = \cos\theta = \frac{\mathrm{A}\cdot\mathrm{B}}{\|\mathrm{A}\|\,\|\mathrm{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\,\sqrt{\sum_{i=1}^{n} B_i^2}} \quad (4)$$

where A_i and B_i are the components of vectors A and B respectively

In this case, the cosine similarity between two documents ranges from 0 to 1, 0 being totally unrelated (no similar terms), and 1 being exactly similar. The closer the similarity score is to 1, the more relevant the documents are to each other

a) Ranking by Unigram Similarity

Like the query, every web document is first tokenized and cleared of stop words, before the TF-IDF values of all its remaining unigrams are calculated and used to form the document's vector. Next, the vector for the query content is aligned to vector of each web document, and the cosine similarity between them is calculated

b) Ranking by Name Similarity

Names of people, organisations and locations are extracted using the Stanford NER from both the query content and the web documents. The TF-IDF value of each name present in every document is calculated and the vectors are formed with these values. The cosine similarity between the query and every web document is computed and compared.

Finally, for every document, its similarity scores for (I) and (II) are then averaged to form the final similarity score, weighted equally, and given its final rank. The list of URLs is then returned to the user as output, ordered by relevance. The process flow of CAWE is given in Fig. 2.

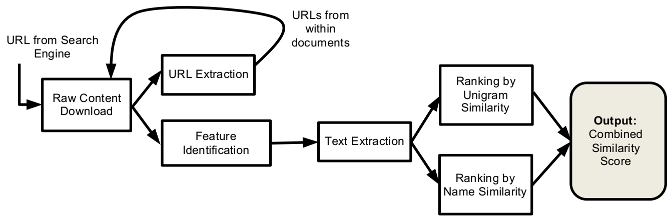


Figure 2. Flowchart of CAWE Processing Process

3) Possible Applications

With more desired information of higher relevance obtained by CAWE, there is potential to harness this information for further useful applications, such as research, profiling and name mining (see explanation below). Two case studies relevant to recent global and regional affairs were conducted to showcase these potential applications (see 3.2 Case Studies below).

Name mining is an extension of CAWE. It extracts names, using the Stanford NER, from the downloaded web documents and determines their importance with the formula given in (5).

$$\text{name importance}_t = \frac{\sum_{i=1}^{N}(\text{TF}_t \times \text{similarity to query}_i)}{\text{N}} \quad (5)$$

where t is the name term, N is the number of documents t appeared in

The names are ranked from the highest to lowest importance to pick out those that are possibly more important and relevant to the query.

B. Engine Testing

In order to test the context-awareness of CAWE, two experiments were conducted. In each experiment, a few documents were picked as query documents and CAWE was run with CAWE-suggested search terms. The engine's results were compared with Google's to test if CAWE could return more relevant results.

In the first experiment, an article on the 2004 Indian Ocean tsunami was input into CAWE as the query (see Appendix A1), with "tsunami" as a vague search term.

In the second experiment, a profile of a man named Albert Lawrence Brooks - whose birth name is Albert Lawrence Einstein - was used as the query (see Appendix B1), with "Albert Einstein" as the search term, in order to test an extreme scenario of famous people overshadowing others who have similar names, as well as to showcase the possibility of context-aware profile research.

CAWE was configured to crawl the web 2 levels deep (i.e. the process of extracting URLs from the web documents and adding them to the processing queue was done once) for each run.

In every run, these few factors from the results were compared:

1. Rank of each engine's (CAWE, Google) top result on the other engine
2. Similarity score of each engine's top result to the query

Relevance of each engine's top result - manually compared with text inspection.

III. Results and Discussion

A. First Run: Tsunami

CAWE completed this run in 2h 30min, returning 27906 ranked URLs (see Appendix A4).

TABLE 1. Summarised results for the first run

	CAWE's Top Result	**Google's Top Result**
Rank in other Engine	Not found (Level 2)	81 (in CAWE)
Unigram Similarity	0.78179	0.69956
Name Similarity	0.53452	0.14302
Similarity to Query	0.65816	0.42129
Remarks	Top ranked document is most relevant	Most relevant document was not found by Google

The highest-ranked web document returned by CAWE is the most relevant one, but was not found in the 32 pages of Google's search results, and had a cosine similarity score higher than that of Google's top result. In fact, CAWE's top result (see Appendix A2) was the original website that the query had been extracted from, while Google's was a Wikipedia page on tsunamis (see Appendix A3), which was not as relevant. While Google performed adequately, with a high unigram similarity score of 0.69956 to the original query, CAWE's top result was still more relevant. The next few websites by CAWE were also found to be more relevant than Google in terms of unigram and name similarity.

B. Second Run: Albert Lawrence Brooks

CAWE completed this run in 4h 31min, returning 40314 ranked URLs (see Appendix B4).

TABLE 2. Summarised results for the second run

	CAWE's Top Result	**Google's Top Result**
Rank in other Engine	Not found (Level 2)	2152 (in CAWE)
Unigram Similarity	0.67088	0.14410
Name Similarity	0.27068	0.02594
Similarity to Query	0.47078	0.08502
Remarks	Top ranked document is most relevant	Most relevant document was not found by Google

As seen from Table 2, CAWE's top result containing information on Albert Brooks (see Appendix B2) could not be found in 42 pages of Google search results, and had a cosine similarity score higher than that of Google's top result, a Wikipedia page on physicist Albert Einstein (see Appendix

B3). In this case, Google's results were completely irrelevant to the query as Google's results were all about the famous physicist. CAWE performed much better as it was able to produce a much more relevant result, with the context provided by the query.

IV. Case Studies

These two case studies showcase the possible applications of CAWE.

A. Counter-Terrorism with Terrorist Profiling

On 2 Dec 2015, two radicalized gunmen, one of whom was Syed Rizwan Farook, opened fire into a crowd during a Christmas party in San Bernardino, California, killing 14. With only a brief profile of him (see Appendix C1), the aims of this case study are to:

1. Collect personal details on Syed beyond the given data, including relationships he might have possibly had with others.
2. Uncover the possible reasons behind the radicalization of Syed, a seemingly normal citizen, and the precursors before his attack.

CAWE was run using the given profile as the query and "Syed Rizwan Farook" as the keywords. It downloaded and ranked a list of 79,490 web documents (see Appendix C2). Name mining was performed, returning 526,048 possibly relevant names (see Appendix C3).

1) Terrorist Profiling

Personal details about Farook were gathered from the top 100 relevant web documents returned by CAWE, and the name mining results contained names of people connected to Farook that were not included in the query. This information allowed us to generate a more comprehensive profile of him that includes his family background, and helped us find people he had possible relations to or who were also involved in terror activities (see Appendix C4).

2) Causes of Radicalization and Precursors before Attack

It was identified that the possible reasons for Farook's radicalization include a violent, fractured family, and interaction with already radicalized individuals. The precursors of the attack were (i) the unusual activity around his house, (ii) his support of radical views, (iii) the transfer of suspicious amounts of money to and from his bank account, and (iv) the visitation of several shooting ranges just before the attack (see Appendix C4 for more details). By using CAWE to study more radicalized individuals, precursors and reasons could be derived to identify them before any terror attacks might be carried out.

B. Disaster Monitoring

The Sabah earthquake on 5 Jun 2015 claimed the lives of 18 mountain-climbers on Mount Kinabalu, 10 of whom were Singaporeans. The objective of this case study is to monitor the impact by monitoring the number of deaths and the number of missing people over time for situation understanding.

CAWE was run with an early news article of the Sabah quake as the query (see Appendix D1), with the search term "Sabah Earthquake". Most relevant articles regarding the quake were output (see Appendix D2), and based on the retrieved information, a timeline of the disaster that focuses on the number of missing and dead people was created.

TABLE 3. Timeline tracking information about the earthquake

Date	Time	Update	Death Toll
5/6/15	7.17am	6.0 magnitude earthquake strikes Ranou, no initial damage reports issued	-
	3.00pm	Discovered 3 rest houses and hostels, Donkey's Ears monument damaged	-
	6.18pm	32 accounted for; 122 missing	-
	8.30pm	118 missing; 2 deaths (1 Malaysian, 1 Singaporean)	2
6/6/15	8.43am	17 missing (6 Malaysians, 1 Chinese, 1 Japanese, 9 Singaporeans)	-
	1.48pm	9 deaths (unidentified)	11
7/6/15	4.01pm	2 missing (2 Singaporeans); 5 deaths (unidentified)	16

As shown in Table 3, CAWE can be used for disaster monitoring by being run periodically to retrieve the most relevant and updated information regarding the disaster. It is more efficient than manually reading pages of search engine results to find information, and also allows for the potential use of natural language processing techniques to extract dates and other important statistics from articles, to easily plot a timeline of events, extrapolate death tolls, etc.

V. Conclusion

In conclusion, CAWE was successful in searching for documents more relevant to the user's query while retaining its context. In our tests, CAWE greatly outperformed Google in web searching, especially in extreme test cases (see 3.1.2) where the user might not have a good understanding of his/her search query to aptly summarize the content into keywords. CAWE is also able to rank and index results of much higher quantity (around 40000 results at level two depth) and perform searches with an unlimited-length query, to give results of much higher relevance.

Preliminary applications of CAWE has yielded results that let us believe the relevant data from CAWE can be used in many practical scenarios, such as profiling, research, humanitarian aid, and intelligence work.

VI. Recommendations for Future Work

With more time, more features can also be explored to further improve CAWE's accuracy and performance. In the Ranker, apart from ranking by unigram and name similarity, other factors can be included, such as ranking by nouns, verbs, or bigrams and trigrams. These can then form the document vectors to be compared. In addition, the unigram and name factors are currently assumed to hold equal importance and are given equal weight. With more training data, a support vector machine can be used to help assign different weights of the factors used to consider document similarity, in order to enhance the accuracy of the Ranker.

Other features that may help widen the scope of the search include: processing multiple languages, and expanding search keywords via synonyms. Features such as spelling error detection and automatic text summarization can be incorporated to improve the user-friendliness, efficiency, and accuracy of CAWE.

CAWE can also be improved to crawl covert Internet channels in order to retrieve more obscure results that may be inaccessible by search engines, and obtain more relevant and more specific results, especially in the circumstances of the case studies detailed above.

ACKNOWLEDGMENT

This project was supported by the DSO National Laboratories. We would like to express our utmost gratitude to our mentor, Dr Terence Tan, for guiding us with his invaluable insight and expertise, and for allowing us the wonderful opportunity to work on this project.

REFERENCES

[1] NetMarketShare. (2015). Desktop Search Engine Market Share. Retrieved from https://www.netmarketshare.com/search-engine-market-share.aspx?qprid=4&qpcustomd=0

[2] Madhu, G., Govardhan A., RajiniKanth T.V. (2011). Intelligent Semantic Web Search Engines: A Brief Survey. In International journal of Web & Semantic Technology (IJWesT) Vol.2, No.1. Retrieved from http://arxiv.org/ftp/arxiv/papers/1102/1102.0831.pdf

[3] IntelliJ IDEA. (n.d.). IntelliJ IDEA. Retrieved from https://www.jetbrains.com/idea/

[4] Wikipedia: Database download. (2015). Wikipedia, the free encyclopedia. Retrieved from https://en.wikipedia.org/wiki/Wikipedia:Database_download

[5] Stanford Named Entity Recognizer (NER). (n.d.). The Stanford Natural Language Processing Group. Retrieved from http://nlp.stanford.edu/software/CRF-NER.shtml

[6] Jsoup Java HTML Parser. (n.d.). Jsoup. Retrieved from http://jsoup.org/

[7] Apache POI. (n.d.). The Apache POI Project. Retrieved from https://poi.apache.org/

[8] Apache PDFBox. (n.d.). PDFBox. Retrieved from https://pdfbox.apache.org/

[9] Lee, M. Zhao, H. (2014). An Improved Cosine Similarity Algorithm Based on Document Similarity. In International Symposium on Fuzzy Systems, Knowledge Discovery and Natural Computation (FSKD 2014). Retrieved from https://books.google.com.sg/books?id=zrGWBQAAQBAJ&source=gbs_navlinks

VII. APPENDICES

Appendix A1: Query (Tsunami)

Source: http://news.bbc.co.uk/2/hi/4136289.stm

The tsunami disaster explained

Tsunamis are caused by earthquakes at sea. Earthquakes happen when the plates that make up the Earth's surface suddenly move against each other.

A tsunami forms when energy from an earthquake vertically jolts the seabed by several metres, displacing hundreds of cubic kilometres of water.

Large waves begin moving through the ocean, away from the earthquake's epicentre.

In deep water, the tsunami moves at great speeds. When it reaches shallow water near coastal areas, the tsunami slows but increases in height.

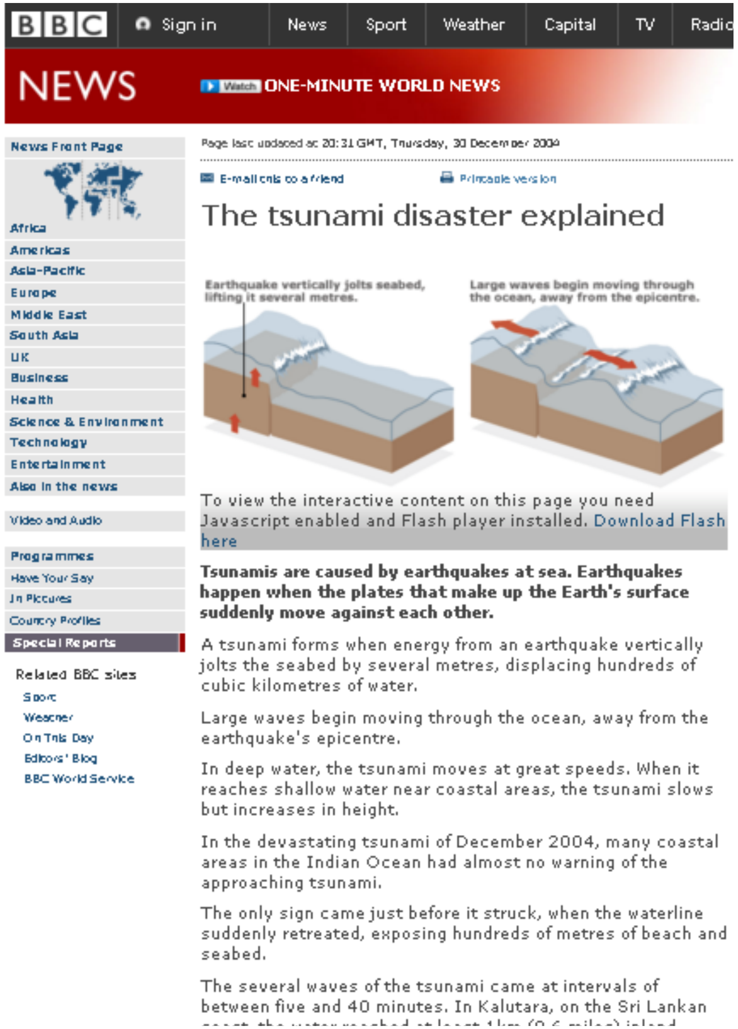
In the devastating tsunami of December 2004, many coastal areas in the Indian Ocean had almost no warning of the approaching tsunami.

The only sign came just before it struck, when the waterline suddenly retreated, exposing hundreds of metres of beach and seabed.

The several waves of the tsunami came at intervals of between five and 40 minutes. In Kalutara, on the Sri Lankan coast, the water reached at least 1km (0.6 miles) inland, causing widespread destruction

Appendix A2: Top Result from CAWE (Tsunami)

Source: http://news.bbc.co.uk/2/hi/4136289.stm


BBC Sign in News Sport Weather Capital TV Radio

NEWS ONE-MINUTE WORLD NEWS

News Front Page
Africa
Americas
Asia-Pacific
Europe
Middle East
South Asia
UK
Business
Health
Science & Environment
Technology
Entertainment
Also in the news
Video and Audio
Programmes
Have Your Say
In Pictures
Country Profiles
Special Reports

Related BBC sites
Sport
Weather
On This Day
Editors' Blog
BBC World Service

Page last updated at 20:31 GMT, Thursday, 30 December 2004

E-mail this to a friend Printable version

The tsunami disaster explained

To view the interactive content on this page you need Javascript enabled and Flash player installed. Download Flash here

Tsunamis are caused by earthquakes at sea. Earthquakes happen when the plates that make up the Earth's surface suddenly move against each other.

A tsunami forms when energy from an earthquake vertically jolts the seabed by several metres, displacing hundreds of cubic kilometres of water.

Large waves begin moving through the ocean, away from the earthquake's epicentre.

In deep water, the tsunami moves at great speeds. When it reaches shallow water near coastal areas, the tsunami slows but increases in height.

In the devastating tsunami of December 2004, many coastal areas in the Indian Ocean had almost no warning of the approaching tsunami.

The only sign came just before it struck, when the waterline suddenly retreated, exposing hundreds of metres of beach and seabed.

The several waves of the tsunami came at intervals of between five and 40 minutes. In Kalutara, on the Sri Lankan

Appendix A3: Top Result from Google (Tsunami)

Source: https://en.wikipedia.org/wiki/Tsunami

As the page is too lengthy, only the text content will be added here.

Tsunami A tsunami (plural: tsunamis or tsunami; from Japanese: 津波, lit. "harbor wave";[1] English pronunciation: /tsuːˈnɑːmi/[2]) , also known as a seismic sea wave, is a series of waves in a water body caused by the displacement of a large volume of water, generally in an ocean or a large lake. Earthquakes, volcanic eruptions and other underwater explosions (including detonations of underwater nuclear devices), landslides, glacier calvings, meteorite impacts and other disturbances above or below water all have the potential to generate a tsunami.[3] Unlike normal ocean waves which are generated by wind or tides which are generated by the gravitational pull of the Moon and Sun, a tsunami is generated by the displacement of water. Tsunami waves do not resemble normal sea waves, because their wavelength is far longer. Rather than appearing as a breaking wave, a tsunami may instead initially resemble a rapidly rising tide, and for this reason they are often referred to as tidal waves, although this usage is not favored by the scientific community because tsunamis are not tidal in nature. Tsunamis generally consist of a series of waves with periods ranging from minutes to hours, arriving in a so-called "wave train".[4] Wave heights of tens of meters can be generated by large events. Although the impact of tsunamis is limited to coastal areas, their destructive power can be enormous and they can affect entire ocean basins; the 2004 Indian Ocean tsunami was among the deadliest natural disasters in human history with at least 230,000 people killed or missing in 14 countries bordering the Indian Ocean. The Greek historian Thucydides suggested in his late-5th century BC History of the Peloponnesian War, that tsunamis were related to submarine earthquakes,[5][6] but the understanding of a tsunami's nature remained slim until the 20th century and much remains unknown. Major areas of current research include trying to determine why some large earthquakes do not generate tsunamis while other smaller ones do; trying to accurately forecast the passage of tsunamis across the oceans; and also to forecast how tsunami waves interact with specific shorelines. Numerous terms are used in the English language to describe waves created in a body of water by the displacement of water; however, none of the terms in frequent use are entirely accurate. The term tsunami, meaning "harbor wave" in literal translation, comes from the Japanese 津波, composed of the two kanji 津 (tsu) meaning "harbor" and 波 (nami), meaning "wave". (For the plural, one can either follow ordinary English practice and add an s, or use an invariable plural as in the Japanese.[7]) There are only a few other languages that have an equivalent native word. In Acehnese, the words are ië beuna[8] or alôn buluëk[9] (depending on the dialect). In Tamil, it is aazhi peralai. On Simeulue island, off the western coast of Sumatra in Indonesia, in Devayan the word is smong, while in Sigulai it is emong.[10] In Singkil (in Aceh province) and surrounding areas, the people use the word gloro/galoro for tsunami.[11][12] In Nias, it is called oloro/galoro[13] and in Ende it is called ae mesi nuka tana lala[14] In Spanish and Portuguese the term used is "Marémoto," literally the "sea moves," and these are often associated with "Terramotos", otherwise known as "Earthquakes." Residents of many coastal areas where Spanish or Portuguese is spoken know that if they feel a terramoto they should seek high ground in case a marémoto occurs. Tsunami are sometimes referred to as tidal waves.[15] This once-popular term derives from the most common appearance of tsunami, which is that of an extraordinarily high tidal bore. Tsunami and tides both produce waves of water that move inland, but in the case of tsunami the inland movement of water may be much greater, giving the impression of an incredibly high and forceful tide. In recent years, the term "tidal wave" has fallen out of favor, especially in the scientific community, because tsunami actually have nothing to do with tides, which are produced by the gravitational pull of the moon and sun rather than the displacement of water. Although the meanings of "tidal" include "resembling"[16] or "having the form or character of"[17] the tides, use of the term tidal wave is discouraged by geologists and oceanographers. The term seismic sea wave also is used to refer to the phenomenon, because the waves most often are generated by seismic activity such as earthquakes.[18] Prior to the rise of the use of the term "tsunami" in English-speaking countries, scientists generally encouraged the use of the term "seismic sea wave" rather than the inaccurate term "tidal wave." However, like "tsunami," "seismic sea wave" is not a completely accurate term, as forces other than earthquakes – including underwater landslides, volcanic eruptions, underwater explosions, land or ice slumping into the ocean, meteorite impacts, or even the weather when the atmospheric pressure changes very rapidly – can generate such waves by displacing water.[19][20] While Japan may have the longest recorded history of tsunamis, the sheer destruction caused by the 2004 Indian Ocean earthquake and tsunami event mark it as the most devastating of its kind in modern times, killing around 230,000 people.[21] The Sumatran region is not unused to tsunamis either, with earthquakes of varying magnitudes regularly occurring off the coast of the island.[22] Tsunamis are an often underestimated hazard in the Mediterranean Sea region and Europe in general. Of historical and current (with regard to risk assumptions) importance are e.g. the 1755 Lisbon earthquake and tsunami (which was caused by the Azores–Gibraltar Transform Fault), the 1783 Calabrian earthquakes, each causing several ten thousand deaths and the 1908 Messina earthquake and tsunami. The tsunami claimed more than 123,000 lives in Sicily and Calabria and is among the most deadly natural disasters in modern Europe. The Storegga Slide in the Norwegian sea and some examples of Tsunamis affecting the British Isles refer to landslide and meteotsunamis predominantly and less to earthquake-induced waves. As early as 426 BC the Greek historian Thucydides inquired in his book History of the Peloponnesian War about the causes of tsunami, and was the first to argue that ocean earthquakes must be the cause.[5][6] "The cause, in my opinion, of this phenomenon must be sought in the earthquake. At the point where its shock has been the most violent the sea is driven back, and suddenly recoiling with redoubled force, causes the inundation. Without an earthquake I do not see how such an accident could happen."[23] The Roman historian Ammianus Marcellinus (Res Gestae 26.10.15-19) described the typical sequence of a tsunami, including an incipient earthquake, the sudden retreat of the sea and a following gigantic wave, after the 365 AD tsunami devastated Alexandria.[24][25] The principal generation mechanism (or cause) of a tsunami is the displacement of a substantial volume of water or perturbation of the sea.[26] This displacement of water is usually attributed to either earthquakes, landslides, volcanic eruptions, glacier calvings or more rarely by meteorites and nuclear tests.[27][28] The waves formed in this way are then sustained by gravity. Tides do not play any part in the generation of tsunamis. Tsunami can be generated when the sea floor abruptly deforms and vertically displaces the overlying water. Tectonic earthquakes are a particular kind of earthquake that are associated with the Earth's crustal deformation; when these earthquakes occur beneath the sea, the water above the deformed area is displaced from its equilibrium position.[29] More specifically, a tsunami can be generated when thrust faults associated with convergent or destructive plate boundaries move abruptly, resulting in water displacement, owing to the vertical component of movement involved. Movement on normal (extensional) faults can also cause displacement of the seabed, but only the largest of such events (typically related to flexure in the outer trench swell) cause enough displacement to give rise to a significant tsunami, such as the 1977 Sumba and 1933 Sanriku events.[30][31] Drawing of tectonic plate boundary before earthquake Overriding plate bulges under strain, causing tectonic uplift. Plate slips, causing subsidence and releasing energy into water. The energy released produces tsunami waves. Tsunamis have a small amplitude (wave height) offshore, and a very long wavelength (often hundreds of kilometres long, whereas normal ocean waves have a wavelength of only 30 or 40 metres),[32] which is why they generally pass unnoticed at sea, forming only a slight swell usually about 300 millimetres (12 in) above the normal sea surface. They grow in height when they reach shallower water, in a wave shoaling process described below. A tsunami can occur in any tidal state and even at low tide can still inundate coastal areas. On April 1, 1946, a magnitude-7.8 (Richter Scale) earthquake occurred near the Aleutian Islands, Alaska. It generated a tsunami which inundated Hilo on the island of Hawai'i with a 14-metre high (46 ft) surge. The area where the earthquake occurred is where the Pacific Ocean floor is subducting (or being pushed downwards) under Alaska. Examples of tsunami originating at locations away from convergent boundaries include Storegga about 8,000 years ago, Grand Banks 1929, Papua New Guinea 1998 (Tappin, 2001). The Grand Banks and Papua New Guinea tsunamis came from earthquakes which destabilised sediments, causing them to flow into the ocean and generate a tsunami. They dissipated before traveling transoceanic distances. The cause of the Storegga sediment failure is unknown. Possibilities include an overloading of the sediments, an earthquake or a release of gas hydrates (methane etc.). The 1960 Valdivia earthquake (Mw 9.5), 1964 Alaska earthquake (Mw 9.2), 2004 Indian Ocean earthquake (Mw 9.2), and 2011 Tōhoku earthquake (Mw9.0) are recent examples of powerful megathrust earthquakes that generated tsunamis (known as teletsunamis) that can cross entire oceans. Smaller (Mw 4.2) earthquakes in Japan can trigger tsunamis (called local and regional tsunamis) that can only devastate nearby coasts, but can do so in only a few minutes. In the 1950s, it was discovered that larger tsunamis than had previously been believed possible could be caused by giant submarine landslides. These rapidly displace large water volumes, as energy transfers to the water at a rate faster than the water can absorb. Their existence was confirmed in 1958, when a giant landslide in Lituya Bay, Alaska, caused the highest wave ever recorded, which had a height of 524 metres (over 1700 feet).[33] The wave did not travel far, as it struck land almost immediately. Two people fishing in the bay were killed, but another boat amazingly managed to ride the wave. Another landslide-tsunami event occurred in 1963 when a massive landslide from Monte Toc entered the Vajont Dam in Italy. The resulting wave surged over the 262 m

(860 ft) high dam by 250 metres (820 ft) and destroyed several towns. Around 2,000 people died.[34][35] Scientists named these waves megatsunamis. Some geologists claim that large landslides from volcanic islands, e.g. Cumbre Vieja on La Palma in the Canary Islands, may be able to generate megatsunamis that can cross oceans, but this is disputed by many others. In general, landslides generate displacements mainly in the shallower parts of the coastline, and there is conjecture about the nature of large landslides that enter water. This has been shown to lead to effect water in enclosed bays and lakes, but a landslide large enough to cause a transoceanic tsunami has not occurred within recorded history. Susceptible locations are believed to be the Big Island of Hawaii, Fogo in the Cape Verde Islands, La Reunion in the Indian Ocean, and Cumbre Vieja on the island of La Palma in the Canary Islands; along with other volcanic ocean islands. This is because large masses of relatively unconsolidated volcanic material occurs on the flanks and in some cases detachment planes are believed to be developing. However, there is growing controversy about how dangerous these slopes actually are.[36] Some meteorological conditions, especially deep depressions such as tropical cyclones, can generate a type of storm surge called a meteotsunami which raises water heights above normal levels, often suddenly at the shoreline.[37] In the case of deep tropical cyclones, this is due to very low atmospheric pressure and inward swirling winds causing an uplifted dome of water to form under and travel in tandem with the storm. When these water domes reach shore, they surge upward in shallows and laterally much as do earthquake-generated tsunamis, typically arriving shortly after landfall of the storm's eye.[38][39] There have been studies of the potential of induction of and at least one actual attempt to create tsunami waves as a tectonic weapon. In World War II, the New Zealand Military Forces initiated Project Seal, which attempted to create small tsunamis with explosives in the area of today's Shakespear Regional Park; the attempt failed.[40] There has been considerable speculation on the possibility of using nuclear weapons to cause tsunamis near to an enemy coastline. Even during World War II consideration of the idea using conventional explosives was explored. Nuclear testing in the Pacific Proving Ground by the United States seemed to generate poor results. Operation Crossroads fired two 20 kilotonnes of TNT (84 TJ) bombs, one in the air and one underwater, above and below the shallow (50 m (160 ft)) waters of the Bikini Atoll lagoon. Fired about 6 km (3.7 mi) from the nearest island, the waves there were no higher than 3–4 m (9.8–13.1 ft) upon reaching the shoreline. Other underwater tests, mainly Hardtack I/Wahoo (deep water) and Hardtack I/Umbrella (shallow water) confirmed the results. Analysis of the effects of shallow and deep underwater explosions indicate that the energy of the explosions doesn't easily generate the kind of deep, all-ocean waveforms which are tsunamis; most of the energy creates steam, causes vertical fountains above the water, and creates compressional waveforms.[41] Tsunamis are hallmarked by permanent large vertical displacements of very large volumes of water which don't occur in explosions. Tsunamis cause damage by two mechanisms: the smashing force of a wall of water travelling at high speed, and the destructive power of a large volume of water draining off the land and carrying a large amount of debris with it, even with waves that do not appear to be large. While everyday wind waves have a wavelength (from crest to crest) of about 100 metres (330 ft) and a height of roughly 2 metres (6.6 ft), a tsunami in the deep ocean has a much larger wavelength of up to 200 kilometres (120 mi). Such a wave travels at well over 800 kilometres per hour (500 mph), but owing to the enormous wavelength the wave oscillation at any given point takes 20 or 30 minutes to complete a cycle and has an amplitude of only about 1 metre (3.3 ft).[42] This makes tsunamis difficult to detect over deep water, where ships are unable to feel their passage. The velocity of a tsunami can be calculated by obtaining the square root of the depth of the water in meters multiplied by the acceleration due to gravity (approximated to 10 m sec2). For example, if the Pacific Ocean is considered to have a depth of 5000 meters, the velocity of a tsunami would be the square root of $\sqrt{5000 \times 10} = \sqrt{50000} = \sim 224$ meters per second (735 feet per second), which equates to a speed of ~806 kilometers per hour or about 500 miles per hour. This formula is the same as used for calculating the velocity of shallow waves, because a tsunami behaves like a shallow wave as it peak to peak value reaches from the floor of the ocean to the surface. The reason for the Japanese name "harbour wave" is that sometimes a village's fishermen would sail out, and encounter no unusual waves while out at sea fishing, and come back to land to find their village devastated by a huge wave. As the tsunami approaches the coast and the waters become shallow, wave shoaling compresses the wave and its speed decreases below 80 kilometres per hour (50 mph). Its wavelength diminishes to less than 20 kilometres (12 mi) and its amplitude grows enormously. Since the wave still has the same very long period, the tsunami may take minutes to reach full height. Except for the very largest tsunamis, the approaching wave does not break, but rather appears like a fast-moving tidal bore.[43] Open bays and coastlines adjacent to very deep water may shape the tsunami further into a step-like wave with a steep-breaking front. When the tsunami's wave peak reaches the shore, the resulting temporary rise in sea level is termed run up. Run up is measured in metres above a reference

sea level.[43] A large tsunami may feature multiple waves arriving over a period of hours, with significant time between the wave crests. The first wave to reach the shore may not have the highest run up.[44] About 80% of tsunamis occur in the Pacific Ocean, but they are possible wherever there are large bodies of water, including lakes. They are caused by earthquakes, landslides, volcanic explosions, glacier calvings, and bolides. All waves have a positive and negative peak, i.e. a ridge and a trough. In the case of a propagating wave like a tsunami, either may be the first to arrive. If the first part to arrive at shore is the ridge, a massive breaking wave or sudden flooding will be the first effect noticed on land. However, if the first part to arrive is a trough, a drawback will occur as the shoreline recedes dramatically, exposing normally submerged areas. Drawback can exceed hundreds of metres, and people unaware of the danger sometimes remain near the shore to satisfy their curiosity or to collect fish from the exposed seabed. A typical wave period for a damaging tsunami is about 12 minutes. This means that if the drawback phase is the first part of the wave to arrive, the sea will recede, with areas well below sea level exposed after 3 minutes. During the next 6 minutes the tsunami wave trough builds into a ridge, and during this time the sea is filled in and destruction occurs on land. During the next 6 minutes, the tsunami wave changes from a ridge to a trough, causing flood waters to drain and drawback to occur again. This may sweep victims and debris some distance from land. The process repeats as the next wave arrives. As with earthquakes, several attempts have been made to set up scales of tsunami intensity or magnitude to allow comparison between different events.[45] The first scales used routinely to measure the intensity of tsunami were the Sieberg-Ambraseys scale, used in the Mediterranean Sea and the Imamura-Iida intensity scale, used in the Pacific Ocean. The latter scale was modified by Soloviev, who calculated the Tsunami intensity I according to the formula where is the average wave height along the nearest coast. This scale, known as the Soloviev-Imamura tsunami intensity scale, is used in the global tsunami catalogues compiled by the NGDC/NOAA[46] and the Novosibirsk Tsunami Laboratory as the main parameter for the size of the tsunami. In 2013, following the intensively studied tsunamis in 2004 and 2011, a new 12 point scale was proposed, the Integrated Tsunami Intensity Scale (ITIS-2012), intended to match as closely as possible to the modified ESI2007 and EMS earthquake intensity scales.[47] The first scale that genuinely calculated a magnitude for a tsunami, rather than an intensity at a particular location was the ML scale proposed by Murty & Loomis based on the potential energy.[45] Difficulties in calculating the potential energy of the tsunami mean that this scale is rarely used. Abe introduced the tsunami magnitude scale , calculated from, where h is the maximum tsunami-wave amplitude (in m) measured by a tide gauge at a distance R from the epicentre, a, b and D are constants used to make the Mt scale match as closely as possible with the moment magnitude scale.[48] Drawbacks can serve as a brief warning. People who observe drawback (many survivors report an accompanying sucking sound), can survive only if they immediately run for high ground or seek the upper floors of nearby buildings. In 2004, ten-year-old Tilly Smith of Surrey, England, was on Maikhao beach in Phuket, Thailand with her parents and sister, and having learned about tsunamis recently in school, told her family that a tsunami might be imminent. Her parents warned others minutes before the wave arrived, saving dozens of lives. She credited her geography teacher, Andrew Kearney. In the 2004 Indian Ocean tsunami drawback was not reported on the African coast or any other east-facing coasts that it reached. This was because the wave moved downwards on the eastern side of the fault line and upwards on the western side. The western pulse hit coastal Africa and other western areas. A tsunami cannot be precisely predicted, even if the magnitude and location of an earthquake is known. Geologists, oceanographers, and seismologists analyse each earthquake and based on many factors may or may not issue a tsunami warning. However, there are some warning signs of an impending tsunami, and automated systems can provide warnings immediately after an earthquake in time to save lives. One of the most successful systems uses bottom pressure sensors, attached to buoys, which constantly monitor the pressure of the overlying water column. Regions with a high tsunami risk typically use tsunami warning systems to warn the population before the wave reaches land. On the west coast of the United States, which is prone to Pacific Ocean tsunami, warning signs indicate evacuation routes. In Japan, the community is well-educated about earthquakes and tsunamis, and along the Japanese shorelines the tsunami warning signs are reminders of the natural hazards together with a network of warning sirens, typically at the top of the cliff of surroundings hills.[49] The Pacific Tsunami Warning System is based in Honolulu, Hawaiʻi. It monitors Pacific Ocean seismic activity. A sufficiently large earthquake magnitude and other information triggers a tsunami warning. While the subduction zones around the Pacific are seismically active, not all earthquakes generate tsunami. Computers assist in analysing the tsunami risk of every earthquake that occurs in the Pacific Ocean and the adjoining land masses. Tsunami hazard sign at Bamfield, British Columbia A tsunami warning sign on a seawall in Kamakura, Japan, 2004 The monument to the victims of the 1946 tsunami at Laupahoehoe, Hawaii Tsunami memorial in Kanyakumari beach A Tsunami hazard sign

(Spanish - English) in Iquique, Chile. Tsunami Evacuation Route signage along U.S. Route 101, in Washington As a direct result of the Indian Ocean tsunami, a re-appraisal of the tsunami threat for all coastal areas is being undertaken by national governments and the United Nations Disaster Mitigation Committee. A tsunami warning system is being installed in the Indian Ocean. Computer models can predict tsunami arrival, usually within minutes of the arrival time. Bottom pressure sensors can relay information in real time. Based on these pressure readings and other seismic information and the seafloor's shape (bathymetry) and coastal topography, the models estimate the amplitude and surge height of the approaching tsunami. All Pacific Rim countries collaborate in the Tsunami Warning System and most regularly practice evacuation and other procedures. In Japan, such preparation is mandatory for government, local authorities, emergency services and the population. Some zoologists hypothesise that some animal species have an ability to sense subsonic Rayleigh waves from an earthquake or a tsunami. If correct, monitoring their behavior could provide advance warning of earthquakes, tsunami etc. However, the evidence is controversial and is not widely accepted. There are unsubstantiated claims about the Lisbon quake that some animals escaped to higher ground, while many other animals in the same areas drowned. The phenomenon was also noted by media sources in Sri Lanka in the 2004 Indian Ocean earthquake.[50][51] It is possible that certain animals (e.g., elephants) may have heard the sounds of the tsunami as it approached the coast. The elephants' reaction was to move away from the approaching noise. By contrast, some humans went to the shore to investigate and many drowned as a result. Along the United States west coast, in addition to sirens, warnings are sent on television and radio via the National Weather Service, using the Emergency Alert System. Kunihiko Shimazaki (University of Tokyo), a member of Earthquake Research committee of The Headquarters for Earthquake Research Promotion of Japanese government, mentioned the plan to public announcement of tsunami attack probability forecast at Japan National Press Club on 12 May 2011. The forecast includes tsunami height, attack area and occurrence probability within 100 years ahead. The forecast would integrate the scientific knowledge of recent interdisciplinarity and aftermath of the 2011 Tōhoku earthquake and tsunami. As the plan, announcement will be available from 2014.[52][53][54] In some tsunami-prone countries earthquake engineering measures have been taken to reduce the damage caused onshore. Japan, where tsunami science and response measures first began following a disaster in 1896, has produced ever-more elaborate countermeasures and response plans.[55] That country has built many tsunami walls of up to 12 metres (39 ft) high to protect populated coastal areas. Other localities have built floodgates of up to 15.5 metres (51 ft) high and channels to redirect the water from incoming tsunami. However, their effectiveness has been questioned, as tsunami often overtop the barriers. The Fukushima Daiichi nuclear disaster was directly triggered by the 2011 Tōhoku earthquake and tsunami, when waves exceeded the height of the plant's sea wall.[56] Iwate Prefecture, which is an area at high risk from tsunami, had tsunami barriers walls totalling 25 kilometres (16 mi) long at coastal towns. The 2011 tsunami toppled more than 50% of the walls and caused catastrophic damage.[57] The Okushiri, Hokkaidō tsunami which struck Okushiri Island of Hokkaidō within two to five minutes of the earthquake on July 12, 1993 created waves as much as 30 metres (100 ft) tall—as high as a 10-story building. The port town of Aonae was completely surrounded by a tsunami wall, but the waves washed right over the wall and destroyed all the wood-framed structures in the area. The wall may have succeeded in slowing down and moderating the height of the tsunami, but it did not prevent major destruction and loss of life.[58]

Appendix A4: CAWE Ranked Results (Tsunami)

Only the top 25 results are shown as the whole output is too large to fit into this document.

CAWE Rank	Word Similarity	Name Similarity	Combined Similarity	Original URL	Item Type	Origin	Level	Search Engine Rank	CSS Selector Tags
1	0.78179	0.53452	0.65816	http://news.bbc.co.uk/1/hi/4136289.stm	HTML	Google	2		p
2	0.60864	0.70711	0.65787	http://www.tsunami2004.net/tag/tsunami-2004/	HTML	Yahoo	2		all
3	0.53378	0.69047	0.61212	http://www.cnn.com/2013/08/20/world/tsunamis-fast-facts/index.html	HTML	Google	2		all
4	0.47485	0.70711	0.59098	http://www.edhelper.com/Tsunami.htm	HTML	Bing	1	268	all
5	0.60724	0.55216	0.57970	http://www.tsunami2004.net/indian-ocean-tsunami-2004/	HTML	Bing	2		all
6	0.42697	0.70711	0.56704	http://www.buzzle.com/articles/tsunamis/	HTML	Bing	2		all
7	0.51993	0.58835	0.55414	http://news.nationalgeographic.com/news/2004/12/1227_041226_tsu	HTML	Bing	2		p
8	0.56274	0.50000	0.53137	http://www.tsunami2004.net/tsunami-2004-lessons-learned/	HTML	Bing	2		p
9	0.51305	0.54772	0.53039	http://news.nationalgeographic.com/news/2004/12/1227_041226_tsu	HTML	Bing	2		all
10	0.55125	0.50000	0.52563	http://www.tsunami2004.net/boxing-day-tsunami-2004-videos/	HTML	Bing	2		all
11	0.53947	0.49029	0.51488	http://wsspc.org/wp-content/uploads/2014/01/TsuCenSigEv_2004Ind	PDF	Google	2		
12	0.57634	0.45291	0.51463	http://www.tsunami2004.net/tsunami-2004-relief/	HTML	Bing	2		all
13	0.63854	0.38490	0.51172	http://www.tsunami2004.net/tsunami-2004-facts/	HTML	Yahoo	1	81	all
14	0.54756	0.47140	0.50948	http://www.tsunami2004.net/tsunami-2004-lessons-learned/	HTML	Bing	2		all
15	0.59251	0.42640	0.50946	http://www.tsunami2004.net/	HTML	Bing	1	69	p
16	0.61153	0.39413	0.50283	http://www.spacedaily.com/news/tectonics-05a.html	HTML	Bing	2		all
17	0.62587	0.37834	0.50211	http://itic.ioc-unesco.org/index.php?option=com_content&view=categ	HTML	Bing	2		all
18	0.53674	0.46291	0.49982	http://asianhistory.about.com/od/asianenvironmentalhistory/p/The-20	HTML	Yahoo	1	76	p
19	0.65337	0.34478	0.49908	http://instructional1.calstatela.edu/sladoch/web_page2/Tsunamis.htm	HTML	Yahoo	1	194	all
20	0.29019	0.70711	0.49865	http://tsun.sscc.ru/tsulab/20041226trt.htm	HTML	Bing	2		all
21	0.28987	0.70711	0.49849	http://www.bbc.co.uk/programmes/p00hs2py	HTML	Google	2		p
22	0.64439	0.34837	0.49638	http://www.bom.gov.au/tsunami/info/	HTML	Yahoo	1	30	p
23	0.48893	0.50000	0.49447	http://english.astroawani.com/malaysia-news/tsunami-baby-eyeing-st	HTML	Bing	2		p
24	0.52072	0.45949	0.49010	http://www.cpp.edu/~marshall/Ind_Oc_Tsunami_Lec.v4.htm	HTML	Bing	1	315	all
25	0.51888	0.45883	0.48886	http://news.nationalgeographic.com/news/2004/12/1228_041228_tsu	HTML	Bing	2		all

Appendix B1: Query (Albert Einstein)

Source: http://www.biography.com/people/albert-brooks-5446

Albert Brooks Biography
Film Actor, Television Actor, Comedian, Director (1947–)
NAME: Albert Brooks
OCCUPATION: Film Actor, Television Actor, Comedian, Director
BIRTH DATE: July 22, 1947 (age 68)
EDUCATION: Carnegie Tech, Beverly Hills High School
PLACE OF BIRTH: Beverly Hills, California
ORIGINALLY: Albert Lawrence Einstein
ZODIAC SIGN: Cancer

For 40 years, Actor and comedian Albert Brooks has worked steadily in Hollywood, appearing both in other people's movies and in his own self-directed projects.
Synopsis
Albert Brooks was born July 22, 1947, in Beverly Hills, CA. He showed talent in stand-up comedy. He made his first major T.V. appearance in 1969. He then worked steadily in T.V., directing several short films. 1976, he took on his first big film role. For the next several years, he took small parts in popular movies until his role Broadcast News. For 40 years, he has worked steadily in Hollywood.
Early Life
Actor and comedian. Born July 22, 1947, in Beverly Hills, California, the youngest son of Thelma Leeds, a singer and actress, and Harry Parke, a radio comedian who went by the stage name of Parkya Karkus. Unable to pass up a good punchline, the two performers named their youngest son Albert Einstein, a name he changed as soon as he became an adult. Albert had two older brothers (including Bob Einstein, a comedian who performs as Super Dave Osborne) and a half-brother on his father's side.
When Brooks was 11 years old, his father died of a heart attack onstage at a Friar's Club roast of Lucille Ball and Desi Arnaz. Harry Parke slumped over into the lap of Milton Berle, whose calls for a doctor were initially met with laughter from the crowd, who thought he was joking. Following this tragedy, Brooks attended Beverly Hills High School, the alma mater of fellow actors Betty White, Rob Reiner and Angelina Jolie. After graduating, he went to Carnegie Tech in Pittsburgh on an acting scholarship, but moved back to Los Angeles a few years later to start his career in show business.
Early Career
Brooks showed major talent on the stand-up comedy circuit. "Everyone knew that Albert was head and shoulders above everybody else, a genius," said Reiner, a friend since high school. "I'm talking people like Chevy Chase, Robin Williams and Billy Crystal basically acknowledging that you couldn't touch this guy. Whenever Albert started in, everyone pretty much backed off and just let him go."
Brooks made his first major television appearance in 1969 on a Dean Martin variety show, doing a bit about a ventriloquist with no talent. He worked steadily in television, directing several short films for the first season of Saturday Night Live and appearing on The Tonight Show with Johnny Carson. In 1976, he took on his first big film role in Martin Scorsese's Taxi Driver, playing a campaign worker in love with a character played by Cybill Shepherd. In 1979, Brooks released Real Life, a mockumentary in which he played an obnoxious filmmaker (named Albert Brooks) who films an American family in an attempt to win an Academy Award. It was the first picture that Brooks wrote, directed and starred in. Though not a smash at the box office, the movie developed a cult following among fans of Brooks' self-mocking humor. He went on to write, direct and star in six more films.
Commercial Success
For the next several years, Brooks moved between small parts in popular movies (such as the boorish husband of Goldie Hawn who dies on his wedding night in Private Benjamin) and writer/director/star roles in little-seen films like Modern Romance. Then in 1985, Brooks wrote, directed and starred in Lost in America, a film about a couple who sell everything, drop out of society and bum around the country in a mobile home. To everyone's surprise (especially Brooks' own), the movie turned out to be a hit.
For his next project, Brooks played a neurotic, insecure television reporter ("Wouldn't this be a great world if insecurity and desperation made us more attractive?") in Broadcast News (1987). The role earned him an Academy Award nomination and cemented his onscreen persona as a neurotic, emotionally complicated guy who has frequently been called the West Coast version of Woody Allen. "It's an interesting world we live in when Arnold Schwarzenegger can kill 115 people in a movie and he's fine," Brooks has said of his reputation. "I drive around a woman's house twice, and I'm neurotic. Go figure."
Later Work
For 40 years, Brooks has worked steadily in Hollywood, appearing both in other people's movies and in his own self-directed projects like Defending Your Life, Mother, The Muse and Looking for Comedy in the Muslim World. In 1997, he married Kimberly Shlain, an artist. The couple has a son and a daughter.
Albert Brooks' career choices have been eclectic. He is a regular guest voice on The Simpsons and voiced a character in the 2003 Pixar hit Finding Nemo. His first novel, 2030: The Real Story of What Happens to America, was published in 2011. An unusual Hollywood career? Maybe, but Brooks plays by his own rules. "If people don't love what you're doing," he has said, "that doesn't mean you're wrong."

Appendix B2: Top Result from CAWE (Albert Einstein)

Source: http://people.famouswhy.com/albert_brooks/

[edit biography]

Albert Brooks

Q: Can I get an image of Albert Brooks? - *by Carlos*

Home / People / Others

Albert Brooks (Albert Lawrence Einstein) was born on Tuesday, July 22, 1947 and is a famous comic.

General appearance:

- 5' 10" (178 cm) height.

Being born on Jul 22, Albert is a Cancer and has a large body.

Albert Brooks had studied at Beverly Hills High School and then Albert attended the Carnegie Tech, Pittsburgh, PA.

Albert dated Carole Mallory and amongst other amorous encounters was Kimberly Shlain.

Albert Brooks Lists

- 20 Most Famous Movies of Sharon Stone (June 28, 2009)
- 20 Most Famous Movies of Lisa Kudrow (June 27, 2009)
- 20 Most Famous Movies of Jeff Bridges (June 24, 2009)
- 20 Most Famous Movies of Jeff Bridges (June 24, 2009)
- 1947 births
- American film actors
- American screenwriters
- American stand-up comedians
- English-language film directors
- American Jews
- Jewish actors
- Jewish comedians
- Jewish American film directors
- Jewish American writers
- Living people
- Actors from Los Angeles
- California

Appendix B3: Top Result from Google (Albert Einstein)

Source: https://en.wikipedia.org/wiki/Albert_Einstein

As the page is too lengthy, only the text content will be added here.

Albert Einstein Albert Einstein (/ˈaɪnstaɪn/;[3] German: [ˈalbɐrt ˈaɪnʃtaɪn] (listen); 14 March 1879 – 18 April 1955) was a German-born theoretical physicist. He developed the general theory of relativity, one of the two pillars of modern physics (alongside quantum mechanics).[2][4]:274 Einstein's work is also known for its influence on the philosophy of science.[5][6] Einstein is best known in popular culture for his mass–energy equivalence formula E = mc2 (which has been dubbed "the world's most famous equation").[7] He received the 1921 Nobel Prize in Physics for his "services to theoretical physics", in particular his discovery of the law of the photoelectric effect, a pivotal step in the evolution of quantum theory.[8] Near the beginning of his career, Einstein thought that Newtonian mechanics was no longer enough to reconcile the laws of classical mechanics with the laws of the electromagnetic field. This led to the development of his special theory of relativity. He realized, however, that the principle of relativity could also be extended to gravitational fields, and with his subsequent theory of gravitation in 1916, he published a paper on general relativity. He continued to deal with problems of statistical mechanics and quantum theory, which led to his explanations of particle theory and the motion of molecules. He also investigated the thermal properties of light which laid the foundation of the photon theory of light. In 1917, Einstein applied the general theory of relativity to model the large-scale structure of the universe.[9][10] He was visiting the United States when Adolf Hitler came to power in 1933 and, being Jewish, did not go back to Germany, where he had been a professor at the Berlin Academy of Sciences. He settled in the U.S., becoming an American citizen in 1940.[11] On the eve of World War II, he endorsed a letter to President Franklin D. Roosevelt alerting him to the potential development of "extremely powerful bombs of a new type" and recommending that the U.S. begin similar research. This eventually led to what would become the Manhattan Project. Einstein supported defending the Allied forces, but largely denounced the idea of using the newly discovered nuclear fission as a weapon. Later, with the British philosopher Bertrand Russell, Einstein signed the Russell–Einstein Manifesto, which highlighted the danger of nuclear weapons. Einstein was affiliated with the Institute for Advanced Study in Princeton, New Jersey, until his death in 1955. Einstein published more than 300 scientific papers along with over 150 non-scientific works.[9][12] On 5 December 2014, universities and archives announced the release of Einstein's papers, comprising more than 30,000 unique documents.[13][14] Einstein's intellectual achievements and originality have made the word "Einstein" synonymous with "genius".[15]   Albert Einstein was born in Ulm, in the Kingdom of Württemberg in the German Empire on 14 March 1879.[16] His parents were Hermann Einstein, a salesman and engineer, and Pauline Koch. In 1880, the family moved to Munich, where his father and his uncle founded Elektrotechnische Fabrik J. Einstein & Cie, a company that manufactured electrical equipment based on direct current.[16] Like Feynman and Teller, Einstein was a late talker.[citation needed] The Einsteins were non-observant Ashkenazi Jews, and Albert attended a Catholic elementary school from the age of 5 for three years. At the age of 8, he was transferred to the Luitpold Gymnasium (now known as the Albert Einstein Gymnasium), where he received advanced primary and secondary school education until he left Germany seven years later.[17] In 1894, his father's company failed: direct current (DC) lost the War of Currents to alternating current (AC). In search of business, the Einstein family moved to Italy, first to Milan and then, a few months later, to Pavia. When the family moved to Pavia, Einstein stayed in Munich to finish his studies at the Luitpold Gymnasium. His father intended for him to pursue electrical engineering, but Einstein clashed with authorities and resented the school's regimen and teaching method. He later wrote that the spirit of learning and creative thought was lost in strict rote learning. At the end of December 1894, he travelled to Italy to join his family in Pavia, convincing the school to let him go by using a doctor's note.[18] It was during his time in Italy that he wrote a short essay with the title "On the Investigation of the State of the Ether in a Magnetic Field.[19][20] In 1895, at the age of 16, Einstein sat the entrance examinations for the Swiss Federal Polytechnic in Zürich (later the Eidgenössische Technische Hochschule, ETH). He failed to reach the required standard in the general part of the examination,[21] but obtained exceptional grades in physics and mathematics.[22] On the advice of the principal of the Polytechnic, he attended the Argovian cantonal school (gymnasium) in Aarau, Switzerland, in 1895–96 to complete his secondary schooling. While lodging with the family of Professor Jost Winteler, he fell in love with Winteler's daughter, Marie. (Albert's sister Maja later married Wintelers' son Paul.)[23] In January 1896, with his father's approval, he renounced his citizenship in the German Kingdom of Württemberg to avoid military service.[24] In September 1896, he passed the Swiss Matura with mostly good grades, including a top grade of 6 in physics and mathematical subjects, on a scale of 1–6.[25] Though only 17, he enrolled in the four-year mathematics and physics teaching diploma program at the Zürich Polytechnic. Marie Winteler moved to Olsberg, Switzerland, for a teaching post. Einstein's future wife, Mileva Marić, also enrolled at the Polytechnic that same year. She was the only woman among the six students in the mathematics and physics section of the teaching diploma course. Over the next few years, Einstein and Marić's friendship developed into romance, and they read books together on extra-curricular physics in which Einstein was taking an increasing interest. In 1900, Einstein was awarded the Zürich Polytechnic teaching diploma, but Marić failed the examination with a poor grade in the mathematics component, theory of functions.[26] There have been claims that Marić collaborated with Einstein on his celebrated 1905 papers,[27][28] but historians of physics who have studied the issue find no evidence that she made any substantive contributions.[29][30][31][32] The discovery and publication in 1987 of an early correspondence between Einstein and Marić revealed that they had had a daughter, called "Lieserl" in their letters, born in early 1902 in Novi Sad where Marić was staying with her parents. Marić returned to Switzerland without the child, whose real name and fate are unknown. Einstein probably never saw his daughter. The contents of his letter to Marić in September 1903 suggest that the girl was either adopted or died of scarlet fever in infancy.[33][34] Einstein and Marić married in January 1903. In May 1904, the couple's first son, Hans Albert Einstein, was born in Bern, Switzerland. Their second son, Eduard, was born in Zürich in July 1910. In 1914, the couple separated; Einstein moved to Berlin and his wife remained in Zürich with their sons. They divorced on 14 February 1919, having lived apart for five years. Eduard, whom his father called "Tete" (for petit), had a breakdown at about age 20 and was diagnosed with schizophrenia. His mother cared for him and he was also committed to asylums for several periods, including full-time after her death. The marriage with Marić does not seem to have been very happy. In letters revealed in 2015, Einstein wrote to his early love, Marie Winteler, about his marriage and his still strong feelings for Marie. In 1910 he wrote to her that "I think of you in heartfelt love every spare minute and am so unhappy as only a man can be" while his wife was pregnant with their second child. Einstein spoke about a "misguided love" and a "missed life" regarding his love for Marie.[35] Einstein married Elsa Löwenthal on 2 June 1919, after having had a relationship with her since 1912. She was a first cousin maternally and a second cousin paternally. In 1933, they emigrated to the United States. In 1935, Elsa Einstein was diagnosed with heart and kidney problems; she died in December 1936.[36] After graduating, Einstein spent almost two frustrating years searching for a teaching post. He acquired Swiss citizenship in February 1901,[37] but was not conscripted for medical reasons. With the help of Marcel Grossmann's father Einstein secured a job in Bern at the Federal Office for Intellectual Property, the patent office,[38][39] as an assistant examiner.[40][41] He evaluated patent applications for a variety of devices including a gravel sorter and an electromechanical typewriter.[41] In 1903, Einstein's position at the Swiss Patent Office became permanent, although he was passed over for promotion until he "fully mastered machine technology".[42]:370 Much of his work at the patent office related to questions about transmission of electric signals and electrical-mechanical synchronization of time, two technical problems that show up conspicuously in the thought experiments that eventually led Einstein to his radical conclusions about the nature of light and the fundamental connection between space and time.[42]:377 With a few friends he had met in Bern, Einstein started a small discussion group, self-mockingly named "The Olympia

Academy", which met regularly to discuss science and philosophy. Their readings included the works of Henri Poincaré, Ernst Mach, and David Hume, which influenced his scientific and philosophical outlook.[43] In 1900, his paper "Folgerungen aus den Capillaritätserscheinungen" ("Conclusions from the Capillarity Phenomena") was published in the prestigious Annalen der Physik.[44][45] On 30 April 1905, Einstein completed his thesis, with Alfred Kleiner, Professor of Experimental Physics, serving as pro-forma advisor. As a result, Einstein was awarded a PhD by the University of Zürich, with his dissertation entitled, "A New Determination of Molecular Dimensions."[1][46] That same year, which has been called Einstein's annus mirabilis (miracle year), he published four groundbreaking papers, on the photoelectric effect, Brownian motion, special relativity, and the equivalence of mass and energy, which were to bring him to the notice of the academic world. By 1908, he was recognized as a leading scientist and was appointed lecturer at the University of Bern. The following year, after giving a lecture on electrodynamics and the relativity principle at the University of Zurich, Alfred Kleiner recommended him to the faculty for a newly created professorship in theoretical physics. Einstein was appointed associate professor in 1909.[47] Einstein became a full professor at the German Charles-Ferdinand University in Prague in April 1911, accepting Austrian citizenship in the Austro-Hungarian empire to do so.[48][49] During his Prague stay Einstein wrote 11 scientific works, 5 of them on radiation mathematics and on quantum theory of the solids. In July 1912 he returned to his alma mater in Zürich. From 1912 until 1914 he was professor of theoretical physics at the ETH Zurich, where he taught analytical mechanics and thermodynamics. He also studied continuum mechanics, the molecular theory of heat, and the problem of gravitation, on which he worked with mathematician and his friend Marcel Grossmann.[50] In 1914, he returned to the German Empire after being appointed director of the Kaiser Wilhelm Institute for Physics (1914–1932)[51] and a professor at the Humboldt University of Berlin, but freed from most teaching obligations. He soon became a member of the Prussian Academy of Sciences, and in 1916 was appointed president of the German Physical Society (1916–1918).[52] Based on calculations Einstein made in 1911, about his new theory of general relativity, light from another star would be bent by the Sun's gravity. In 1919 that prediction was confirmed by Sir Arthur Eddington during the solar eclipse of 29 May 1919. Those observations were published in the international media, making Einstein world famous. On 7 November 1919, the leading British newspaper The Times printed a banner headline that read: "Revolution in Science – New Theory of the Universe – Newtonian Ideas Overthrown".[53] In 1920, he became Foreign Member of the Royal Netherlands Academy of Arts and Sciences.[54] In 1921, Einstein was awarded the Nobel Prize in Physics for his explanation of the photoelectric effect, as relativity was considered still somewhat controversial. Einstein was elected a Foreign Member of the Royal Society (ForMemRS) in 1921.[2] He also received the Copley Medal from the Royal Society in 1925.[2] Einstein visited New York City for the first time on 2 April 1921, where he received an official welcome by Mayor John Francis Hylan, followed by three weeks of lectures and receptions. He went on to deliver several lectures at Columbia University and Princeton University, and in Washington he accompanied representatives of the National Academy of Science on a visit to the White House. On his return to Europe he was the guest of the British statesman and philosopher Viscount Haldane in London, where he met several renowned scientific, intellectual and political figures, and delivered a lecture at King's College.[55] [56] He also published an essay, "My First Impression of the U.S.A.," in July 1921, in which he tried briefly to describe some characteristics of Americans, much as had Alexis de Tocqueville, who published his own impressions in Democracy in America (1835).[57] For some of his observations, Einstein was clearly surprised: "What strikes a visitor is the joyous, positive attitude to life . . . The American is friendly, self-confident, optimistic, and without envy."[58]:20 In 1922, his travels took him to Asia and later to Palestine, as part of a six-month excursion and speaking tour, as he visited Singapore, Ceylon and Japan, where he gave a series of lectures to thousands of Japanese. After his first public lecture, he met the emperor and empress at the Imperial Palace, where thousands came to watch. In a letter to his sons, Einstein described his impression of the Japanese as being modest, intelligent, considerate, and having a true feel for art.[59] Because of Einstein's travels to the Far East, he was unable to personally accept the Nobel Prize for Physics at the Stockholm award ceremony in December 1922. In his place, the banquet speech was held by a German diplomat, who praised Einstein not only as a scientist but also as an international peacemaker and activist.[60] On his return voyage, he visited Palestine for 12 days in what would become his only visit to that region. Einstein was greeted as if he were a head of state, rather than a physicist, which included a cannon salute upon arriving at the home of the British high commissioner, Sir Herbert Samuel. During one reception, the building was stormed by people who wanted to see and hear him. In Einstein's talk to the audience, he expressed happiness that the Jewish people were beginning to be recognized as a force in the world.[61] In December 1930, Einstein visited America for the second time, originally intended as a two-month working visit as a research fellow at the California Institute of Technology. After the national attention he received during his first trip to the U.S., he and his arrangers aimed to protect his privacy. Although swamped with telegrams and invitations to receive awards or speak publicly, he declined them all.[62] After arriving in New York City, Einstein was taken to various places and events, including Chinatown, a lunch with the editors of the New York Times, and a performance of Carmen at the Metropolitan Opera, where he was cheered by the audience on his arrival. During the days following, he was given the keys to the city by Mayor Jimmy Walker and met the president of Columbia University, who described Einstein as "the ruling monarch of the mind."[63] Harry Emerson Fosdick, pastor at New York's Riverside Church, gave Einstein a tour of the church and showed him a full-size statue that the church made of Einstein, standing at the entrance.[63] Also during his stay in New York, he joined a crowd of 15,000 people at Madison Square Garden during a Hanukkah celebration.[63] Einstein next traveled to California where he met Caltech president and Nobel laureate, Robert A. Millikan. His friendship with Millikan was "awkward", as Millikan "had a penchant for patriotic militarism," where Einstein was a pronounced pacifist.[64] During an address to Caltech's students, Einstein noted that science was often inclined to do more harm than good.[65] This aversion to war also led Einstein to befriend author Upton Sinclair and film star Charlie Chaplin, both noted for their pacifism. Carl Laemmle, head of Universal Studios, gave Einstein a tour of his studio and introduced him to Chaplin. They had an instant rapport, with Chaplin inviting Einstein and his wife, Elsa, to his home for dinner. Chaplin said Einstein's outward persona, calm and gentle, seemed to conceal a "highly emotional temperament," from which came his "extraordinary intellectual energy."[66]:320 Chaplin also remembers Elsa telling him about the time Einstein conceived his theory of relativity. During breakfast one morning, he seemed lost in thought and ignored his food. She asked him if something was bothering him. He sat down at his piano and started playing. He continued playing and writing notes for half an hour, then went upstairs to his study, where he remained for two weeks, with Elsa bringing up his food. At the end of the two weeks he came downstairs with two sheets of paper bearing his theory.[66]:320 Chaplin's film, City Lights, was to premier a few days later in Hollywood, and Chaplin invited Einstein and Elsa to join him as his special guests. Walter Isaacson, Einstein's biographer, described this as "one of the most memorable scenes in the new era of celebrity." Einstein and Chaplin arrived together, in black tie, with Elsa joining them, "beaming." The audience applauded as they entered the theater.[65] Chaplin visited Einstein at his home on a later trip to Berlin, and recalled his "modest little flat" and the piano at which he had begun writing his theory. Chaplin speculated that it was "possibly used as kindling wood by the Nazis."[66]:322 In February 1933 while on a visit to the United States, Einstein knew he could not return to Germany with the rise to power of the Nazis under Germany's new chancellor, Adolf Hitler.[67][68] While at American universities in early 1933, he undertook his third two-month visiting professorship at the California Institute of Technology in Pasadena. He and his wife Elsa returned to Belgium by ship in March, and during the trip they learned that their cottage was raided by the Nazis and his personal sailboat confiscated. Upon landing in Antwerp on 28 March, he immediately went to the German consulate and turned in his passport, formally renouncing his German citizenship.[69] A few years later, the Nazis sold his boat and turned his cottage into an Aryan youth camp.[70] In April 1933, he also discovered that the new German government had passed laws barring Jews from holding any official positions, including teaching at universities.[69] Historian Gerald Holton describes how, with "virtually no audible protest being raised by their colleagues," thousands of Jewish scientists were suddenly forced to give up their university positions and their names were removed from the rolls of institutions where they were employed.[58] A month later, Einstein's works were among those targeted by Nazi book burnings, with Nazi propaganda minister Joseph Goebbels proclaiming, "Jewish intellectualism is dead."[69] One German magazine included him in a list of enemies of the German regime with the phrase, "not yet hanged", offering a $5,000 bounty on his head.[69][71] In a subsequent letter to physicist and friend Max Born, who had already emigrated from Germany to England, Einstein wrote, "... I must confess that the degree of their brutality and cowardice came as something of a surprise."[69] After moving to the U.S., he described the book burnings as a "spontaneous emotional outburst" by those who "shun popular enlightenment," and "more than anything else in the world, fear the influence of men of intellectual independence."[72] Einstein was now without a permanent home, unsure where he would live and work, and equally worried about the fate of countless other scientists still in Germany. He rented a house in De Haan, Belgium, where he lived for a few months. In late July 1933, he went to England for about six weeks at the personal invitation of British naval officer Commander Oliver Locker-Lampson, who had become friends with Einstein in the preceding years. To protect Einstein, Locker-Lampson secretly had two assistants watch over him at his secluded cottage outside of London, with the press publishing a photo of them guarding Einstein.[73] Locker-Lampson took Einstein to meet Winston Churchill at his home, and later, Austen Chamberlain and former Prime Minister Lloyd George.[74] Einstein asked them to help bring Jewish scientists out of Germany. British historian Martin Gilbert notes that Churchill responded immediately, and sent his friend, physicist Frederick Lindemann to Germany to seek out Jewish scientists and place them in British universities.[75] Churchill later observed that as a result of Germany having driven the Jews out, they had lowered their "technical standards" and put the Allies' technology ahead of theirs.[75] Einstein later contacted leaders of other nations, including Turkey's Prime Minister, İsmet İnönü, to whom he wrote in September 1933 requesting placement of unemployed German-Jewish scientists. As a result of Einstein's letter, Jewish invitees to Turkey eventually totaled over "1,000 saved individuals."[76] Locker-Lampson also submitted a bill to parliament to extend British citizenship to Einstein, during which period Einstein made a number of public appearances describing the crisis brewing in Europe. The bill failed to become law, however, and Einstein then accepted an earlier offer from the Princeton Institute for Advanced Study, in the U.S., to become a resident scholar.[77] In October 1933 Einstein returned to the U.S. and took up a position at the Institute for Advanced Study (in Princeton, New Jersey),[77][78] noted for having become a refuge for scientists fleeing Nazi Germany.[79] At the time, most American universities, including Harvard, Princeton and Yale, had minimal or no Jewish faculty or students, as a result of their Jewish quota which lasted until the late 1940s.[79] Einstein was still undecided on his future. He had offers from several European universities, including Oxford where he stayed for three short periods between May 1931 and June 1933,[80][81] but in 1935 he arrived at the decision to remain permanently in the United States and apply for citizenship.[77][82] Einstein's affiliation with the Institute for Advanced Study would last until his death in 1955.[83] He was one of the four first selected (two of the others being John von Neumann and Kurt Gödel) at the new Institute, where he soon developed a close friendship with Gödel. The two would take long walks together discussing their work. Bruria Kaufman, his assistant, later became a physicist. During this period, Einstein tried to develop a unified field theory and to refute the accepted interpretation of quantum physics, both unsuccessfully. In 1939, a group of Hungarian scientists that included émigré physicist Leó Szilárd attempted to alert Washington of ongoing Nazi atomic bomb research. The group's warnings were discounted.[84] Einstein and Szilárd, along with other refugees such as Edward Teller and Eugene Wigner, "regarded it as their responsibility to alert Americans to the possibility that German scientists might win the race to build an atomic bomb, and to warn that Hitler would be more than willing to resort to such a weapon."[85][86] To make certain the U.S. was aware of the danger, in July 1939, a few months before the beginning of World War II in Europe, Szilárd and Wigner visited Einstein to explain the possibility of atomic bombs, which Einstein, a pacifist, said he had never considered.[87] He was asked to lend his support by writing a letter, with Szilárd, to President Roosevelt, recommending the U.S. pay attention and engage in its own nuclear weapons research. A secret German facility, apparently the largest of the Third Reich, covering 75 acres in an underground complex, was being re-excavated in Austria in December 2014 and may have been planned for use in nuclear research and development.[88] The letter is believed to be "arguably the key stimulus for the U.S. adoption of serious investigations into nuclear weapons on the eve of the U.S. entry into World War II".[89] In addition to the letter, Einstein used his connections with the Belgian Royal Family[90] and the Belgian queen mother[84] to get access with a personal envoy to the White House's Oval Office.[84] President Roosevelt could not take the risk of allowing Hitler to possess atomic bombs first. As a result of Einstein's letter and his meetings with Roosevelt, the U.S. entered the "race" to develop the bomb, drawing on its "immense material, financial, and scientific resources" to initiate the Manhattan Project. It became the only country to successfully develop an atomic bomb during World War II. For Einstein, "war was a disease ... [and] he called for resistance to war." By signing the letter to Roosevelt he went against his pacifist principles.[91] In 1954, a year before his death, Einstein said to his old friend, Linus Pauling, "I made one great mistake in my life—when I signed the letter to President Roosevelt recommending that atom bombs be made; but there was some justification—the danger that the Germans would make them ..."[92] Einstein became an American citizen in 1940. Not long after settling into his career at the Institute for Advanced Study (in Princeton, New Jersey), he expressed his appreciation of the meritocracy in American culture when compared to Europe. He recognized the "right of individuals to say and think what they pleased", without social barriers, and as a result, individuals were encouraged, he said, to be more creative, a trait he valued from his own early education.[93] Einstein was a passionate, committed antiracist and joined National Association for the Advancement of Colored People (NAACP) in Princeton, where he campaigned for the civil rights of African Americans. He considered racism America's "worst disease,"[71] seeing it as "handed down from one generation to the next."[94] As part of his involvement, he corresponded with civil rights activist W. E. B. Du Bois and was prepared to testify on his behalf during his trial in 1951.[95]:565 When Einstein offered to be a character witness for Du Bois, the judge decided to drop the case.[96] In 1946 Einstein visited Lincoln University in Pennsylvania where he was awarded an honorary degree. Lincoln was the first university in the United States to grant college degrees to blacks, including Langston Hughes and Thurgood Marshall. To its students, Einstein gave a speech about racism in America, adding, "I do not intend to be quiet about it."[97] A resident of Princeton recalls that Einstein had once paid the college tuition for a black student,[96] and black physicist Sylvester James Gates states that Einstein had been one of his early science heroes, later finding out about Einstein's support for civil rights.[96] Einstein was a figurehead leader in helping establish the Hebrew University of Jerusalem, which opened in 1925, and was among its first Board of Governors. Earlier, in 1921, he was asked by the biochemist and president of the World Zionist Organization, Chaim Weizmann, to help raise funds for the planned university.[98] He also submitted various suggestions as to its initial programs. Among those, he advised first creating an Institute of Agriculture in order to settle the undeveloped land. That should be followed, he suggested, by a Chemical Institute and an Institute of Microbiology, to fight the various ongoing epidemics such as malaria, which he called an "evil" that was undermining a third of the country's development.[99]:161 Establishing an Oriental Studies Institute, to include language courses given in both Hebrew and Arabic, for scientific exploration of the country and its historical monuments, was also important.[99]:158 Chaim Weizmann later became Israel's first president. Upon his death while in office in November 1952 and at the urging of Ezriel Carlebach, Prime Minister David Ben-Gurion offered Einstein the position of President of Israel, a mostly ceremonial post.[100][101] The offer was presented by Israel's ambassador in Washington, Abba Eban, who explained that the offer "embodies the deepest respect which the Jewish people can repose in any of its sons".[102] Einstein declined, and wrote in his response that he was "deeply moved", and "at once saddened and ashamed" that he could not accept it.[102] If I were not a physicist, I would probably be a musician. I often think in music. I live my daydreams in music. I see my life in terms of music... I get most joy in life out of music. Einstein developed an appreciation of music at an early age. His mother played the piano reasonably well and wanted her son to learn the violin, not only to instill in him a love of music but also to help him assimilate into German culture. According to conductor Leon Botstein, Einstein is said to have begun playing when he was 5, although he did not enjoy it at that age.[105] When he turned 13 he discovered the violin sonatas of Mozart, whereupon "Einstein fell in love" with Mozart's music and studied music more willingly. He taught himself to play without "ever practicing systematically", he said, deciding that "love is a better teacher than a sense of duty."[105] At age 17, he was heard by a school examiner in Aarau as he played Beethoven's violin sonatas, the examiner stating afterward that his playing was "remarkable and revealing of 'great insight'." What struck the examiner, writes Botstein, was that Einstein "displayed a deep love of the music, a quality that was and remains in short supply. Music possessed an unusual meaning for this student."[105] Music took on a pivotal and permanent role in Einstein's life from that period on. Although the idea of becoming a professional himself was not on his mind at any time, among those with whom Einstein played chamber music were a few professionals, and he performed for private audiences and friends. Chamber music had also become a regular part of his social life while living in Bern, Zürich, and Berlin, where he played with Max Planck and his son, among others. He is sometimes erroneously credited as the editor of the 1937 edition of the Köchel catalogue of Mozart's work; that edition was actually prepared by Alfred Einstein. In 1931, while engaged in research at the California Institute of Technology, he visited the Zoellner family conservatory in Los Angeles, where he played some of Beethoven and Mozart's works with members of the Zoellner Quartet.[106][107] Near the end of his life, when the young Juilliard Quartet visited him in Princeton, he played his violin with them, and the quartet was "impressed by Einstein's level of coordination and intonation."[105] Einstein's political view was in favor of socialism and critical of capitalism, which he detailed in his essays such as "Why Socialism?".[108][109] Einstein offered and was called on to give judgments and opinions on matters often unrelated to theoretical physics or mathematics.[77] He strongly advocated the idea of a democratic global government that would check the power of nation-states in the framework of a world federation.[110] Einstein's views about religious belief have been collected from interviews and original writings. He called himself an agnostic, while disassociating himself from the label atheist.[111] He said he believed in the "pantheistic" God of Baruch Spinoza, but not in a personal god, a belief he criticized.[112][113] Einstein once wrote: "I do not believe in a personal God and I have never denied this but expressed it clearly".[114] On 17 April 1955, Albert Einstein experienced internal bleeding caused by the rupture of an abdominal aortic aneurysm, which had previously been reinforced surgically by Rudolph Nissen in 1948.[115] He took the draft of a speech he was preparing for a television appearance commemorating the State of Israel's seventh anniversary with him to the hospital, but he did not live long enough to complete it.[116] Einstein refused surgery, saying: "I want to go when I want. It is tasteless to prolong life artificially. I have done my share, it is time to go. I will do it elegantly."[117] He died in Princeton Hospital early the next morning at the age of 76, having continued to work until near the end. During the autopsy, the pathologist of Princeton Hospital, Thomas Stoltz Harvey, removed Einstein's brain for preservation without the permission of his family, in the hope that the neuroscience of the future would be able to discover what made Einstein so intelligent.[118] Einstein's remains were cremated and his ashes were scattered at an undisclosed location.[119][120] In his lecture at Einstein's memorial, nuclear physicist Robert Oppenheimer summarized his impression of him as a person: "He was almost wholly without sophistication and wholly without worldliness ... There was always with him a wonderful purity at once childlike and profoundly stubborn."[121] Throughout his life, Einstein published hundreds of books and articles.[12][16] He published more than 300 scientific papers and 150 non-scientific ones.[9][12] On 5 December 2014, universities and archives announced the release of Einstein's papers, comprising more than 30,000 unique documents.[13][14] Einstein's intellectual achievements and originality have made the word "Einstein" synonymous with "genius".[15] In addition to the work he did by himself he also collaborated with other scientists on additional projects including the Bose–Einstein statistics, the Einstein refrigerator and others.[122] The Annus Mirabilis papers are four articles pertaining to the photoelectric effect (which gave rise to quantum theory), Brownian motion, the special theory of relativity, and E = mc2 that Albert Einstein published in the Annalen der Physik scientific journal in 1905. These four works contributed substantially to the foundation of modern physics and changed views on space, time, and matter. The four papers are: Albert Einstein's first paper[127] submitted in 1900 to Annalen der Physik was on capillary attraction. It was published in 1901 with the title "Folgerungen aus den Capillaritätserscheinungen", which translates as "Conclusions from the capillarity phenomena". Two papers he published in 1902–1903 (thermodynamics) attempted to interpret atomic phenomena from a statistical point of view. These papers were the foundation for the 1905 paper on Brownian motion, which showed that Brownian movement can be construed as firm evidence that molecules exist. His research in 1903 and 1904 was mainly concerned with the effect of finite atomic size on diffusion phenomena.[127] He articulated the principle of relativity. This was understood by Hermann Minkowski to be a generalization of rotational invariance from space to space-time. Other principles postulated by Einstein and later vindicated are the principle of equivalence and the principle of adiabatic invariance of the quantum number. Einstein's "Zur Elektrodynamik bewegter Körper" ("On the Electrodynamics of Moving Bodies") was received on 30 June 1905 and published 26 September of that same year. It reconciles Maxwell's equations for electricity and magnetism with the laws of mechanics, by introducing major changes to mechanics close to the speed of light. This later became known as Einstein's special theory of relativity. Consequences of this include the time-space frame of a moving body appearing to slow down and contract (in the direction of motion) when measured in the frame of the observer. This paper also argued that the idea of a luminiferous aether—one of the leading theoretical entities in physics at the time—was superfluous.[128] In his paper on mass–energy equivalence, Einstein produced E = mc2 from his special relativity equations.[129] Einstein's 1905 work on relativity remained controversial for many years, but was accepted by leading physicists, starting with Max Planck.[130][131] In a 1905 paper,[132] Einstein postulated that light itself consists of localized particles (quanta). Einstein's light quanta were nearly universally rejected by all physicists, including Max Planck and Niels Bohr. This idea only became universally accepted in 1919, with Robert Millikan's detailed experiments on the photoelectric effect, and with the measurement of Compton scattering. Einstein concluded that each wave of frequency f is associated with a collection of photons with energy hf each, where h is Planck's constant. He does not say much more, because he is not sure how the particles are related to the wave. But he does suggest that this idea would explain certain experimental results, notably the photoelectric effect.[132] In 1907, Einstein proposed a model of matter where each atom in a lattice structure is an independent harmonic oscillator. In the Einstein model, each atom oscillates independently—a series of equally spaced quantized states for each oscillator. Einstein was aware that getting the frequency of the actual oscillations would be different, but he nevertheless proposed this theory because it was a particularly clear demonstration that quantum mechanics could solve the specific heat problem in classical mechanics. Peter Debye refined this model.[133] Throughout the 1910s, quantum mechanics expanded in scope to cover many different systems. After Ernest Rutherford discovered the nucleus and proposed that electrons orbit like planets, Niels Bohr was able to show that the same quantum mechanical postulates introduced by Planck and developed by Einstein would explain the discrete motion of electrons in atoms, and the periodic table of the elements. Einstein contributed to these developments by linking them with the 1898 arguments Wilhelm Wien had made. Wien had shown that the hypothesis of adiabatic invariance of a thermal equilibrium state allows all the blackbody curves at different temperature to be derived from one another by a simple shifting process. Einstein noted in 1911 that the same adiabatic principle shows that the quantity which is quantized in any mechanical motion must be an adiabatic invariant. Arnold Sommerfeld identified this adiabatic invariant as the action variable of classical mechanics. Although the patent office promoted Einstein to Technical Examiner Second Class in 1906, he had not given up on academia. In 1908, he became a Privatdozent at the University of Bern.[134] In "über die Entwicklung unserer Anschauungen über das Wesen und die Konstitution der Strahlung" ("The Development of our Views on the Composition and Essence of Radiation"), on the quantization of light, and in an earlier 1909 paper, Einstein showed that Max Planck's energy quanta must have well-defined momenta and act in some respects as independent, point-like particles. This paper introduced the photon concept (although the name photon was introduced later by Gilbert N. Lewis in 1926) and inspired the notion of wave–particle duality in quantum mechanics. Einstein saw this wave-particle duality in radiation as concrete evidence for his conviction that physics needed a new, unified foundation. Einstein returned to the problem of thermodynamic fluctuations, giving a treatment of the density variations in a fluid at its critical point. Ordinarily the density fluctuations are controlled by the second derivative of the free energy with respect to the density. At the critical point, this derivative is zero, leading to large fluctuations. The effect of density fluctuations is that light of all wavelengths is scattered, making the fluid look milky white. Einstein relates this to Rayleigh scattering, which is what happens when the fluctuation size is much smaller than the wavelength, and which explains why the sky is blue.[135] Einstein quantitatively derived critical opalescence from a treatment of density fluctuations, and demonstrated how both the effect and Rayleigh scattering originate from the atomistic constitution of matter. In a series of works completed from 1911 to 1913, Planck reformulated his 1900 quantum theory and introduced the idea of zero-point energy in his "second quantum theory." Soon, this idea attracted the attention of Albert Einstein and his assistant Otto Stern. Assuming the energy of rotating diatomic molecules contains zero-point energy, they then compared the theoretical specific heat of hydrogen gas with the experimental data. The numbers matched nicely. However, after publishing the findings, they promptly withdrew their support, because they no longer had confidence in the correctness of the idea of zero-point energy.[136] General relativity (GR) is a theory of gravitation that was developed by Albert Einstein between 1907 and 1915. According to general relativity, the observed gravitational attraction between masses results from the warping of space and time by those masses. General relativity has developed into an essential tool in modern astrophysics. It provides the foundation for the current understanding of black holes, regions of space where gravitational attraction is so strong that not even light can escape. As Albert Einstein later said, the reason for the development of general relativity was that the preference of inertial motions within special relativity was unsatisfactory, while a theory which from the outset prefers no state of motion (even accelerated ones) should appear more satisfactory.[137] Consequently, in 1907 he published an article on acceleration under special relativity. In that article titled "On the Relativity Principle and the Conclusions Drawn from It", he argued that free fall is really inertial motion, and that for a free-falling observer the rules of special relativity must apply. This argument is called the equivalence principle. In the same article, Einstein also predicted the phenomena of gravitational time dilation, gravitational red shift and deflection of light.[138][139] In 1911, Einstein published another article "On the Influence of Gravitation on the Propagation of Light" expanding on the 1907 article, in which he estimated the amount of deflection of light by massive bodies. Thus, the theoretical prediction of general relativity can for the first time be tested experimentally.[140] While developing general relativity, Einstein became confused about the gauge invariance in the theory. He formulated an argument that led him to conclude that a general relativistic field theory is impossible. He gave up looking for fully generally covariant tensor equations, and searched for equations that would be invariant under general linear transformations only. In June 1913, the Entwurf ("draft") theory was the result of these investigations. As its name suggests, it was a sketch of a theory, less elegant and more difficult than general relativity, with the equations of motion supplemented by additional gauge fixing conditions. After more than two years of intensive work, Einstein realized that the hole argument was mistaken[141] and abandoned the theory in November 1915. In 1917, Einstein applied the general theory of relativity to the structure of the universe as a whole.[142] He discovered that the general field equations predicted a universe that was dynamic, either contracting or expanding. As observational evidence for a dynamic universe was not known at the time, Einstein introduced a new term, the cosmological constant, to the field equations, in order to allow the theory to predict a static universe. The modified field equations predicted a static universe of closed curvature, in accordance with Einstein's understanding of Mach's principle in these years.[142][143] Following the discovery of the recession of the nebulae by Edwin Hubble in 1929, Einstein abandoned his static model of the universe, and proposed two dynamic models of the cosmos, the Friedman-Einstein model of 1931[144][145] and the Einstein-deSitter model of 1932.[146] In each of these models, Einstein discarded the cosmological constant, claiming that it was "in any case theoretically unsatisfactory".[144][145][147] In many Einstein biographies, it is claimed that Einstein referred to the cosmological constant in later years as his "biggest blunder". The astrophysicist Mario Livio has recently cast doubt on this claim, suggesting that it may be exaggerated.[148] In late 2013, a team led by the Irish physicist Cormac O'Raifeartaigh discovered evidence that, shortly after learning of Hubble's observations of the recession of the nebulae, Einstein considered a steady-state model of the universe.[149] In a hitherto overlooked manuscript, apparently written in early 1931, Einstein explored a model of the expanding universe in which the density of matter remains constant due to a continuous creation of matter, a process he associated with the cosmological constant.[150][151] As he stated in the paper, "In what follows, I would like to draw attention to a solution to equation (1) that can account for Hubbel's [sic] facts, and in which the density is constant over time"..."If one considers a physically bounded volume, particles of matter will be continually leaving it. For the density to remain constant, new particles of matter must be continually formed in the volume from space." It thus appears that Einstein considered a Steady State model of the expanding universe many years before Hoyle, Bondi and Gold.[152][153] However, Einstein's steady-state model contained a fundamental flaw and he quickly abandoned the idea.[150][151][154] Einstein was displeased with quantum theory and quantum mechanics (the very theory he helped create), despite its acceptance by other physicists, stating that God "is not playing at dice."[155] Einstein continued to maintain his disbelief in the theory, and attempted unsuccessfully to disprove it until he died at the age of 76.[156] In 1917, at the height of his work on relativity, Einstein published an article in Physikalische Zeitschrift that proposed the possibility of stimulated emission, the physical process that makes possible the maser and the laser.[157] This article showed that the statistics of absorption and emission of light would only be consistent with Planck's distribution law if the emission of light into a mode with n photons would be enhanced statistically compared to the emission of light into an empty mode. This paper was enormously influential in the later development of quantum mechanics, because it was the first paper to show that the statistics of atomic transitions had simple laws. Einstein discovered Louis de Broglie's work, and supported his ideas, which were received skeptically at first. In another major paper from this era, Einstein gave a wave equation for de Broglie waves, which Einstein suggested was the Hamilton–Jacobi equation of mechanics. This paper would inspire Schrödinger's work of 1926. In 1924, Einstein received a description of a statistical model from Indian physicist Satyendra Nath Bose, based on a counting method that assumed that light could be understood as a gas of indistinguishable particles. Einstein noted that Bose's statistics applied to some atoms as well as to the proposed light particles, and submitted his translation of Bose's paper to the Zeitschrift für Physik. Einstein also published his own articles describing the model and its implications, among them the Bose–Einstein condensate phenomenon that some particulates should appear at very low temperatures.[158] It was not until 1995 that the first such condensate was produced experimentally by Eric Allin Cornell and Carl Wieman using ultra-cooling equipment built at the NIST–JILA laboratory at the University of Colorado at Boulder.[159] Bose–Einstein statistics are now used to describe the behaviors of any assembly of bosons. Einstein's sketches for this project may be seen in the Einstein Archive in the library of the Leiden University.[122] General relativity includes a dynamical spacetime, so it is difficult to see how to identify the conserved energy and momentum. Noether's theorem allows these quantities to be determined from a Lagrangian with translation invariance, but general covariance makes translation invariance into something of a gauge symmetry. The energy and momentum derived within general relativity by Noether's presecriptions do not make a real tensor for this reason. Einstein argued that this is true for fundamental reasons, because the gravitational field could be made to vanish by a choice of coordinates. He maintained that the non-covariant energy momentum pseudotensor was in fact the best description of the energy momentum distribution in a gravitational field. This approach has been echoed by Lev Landau and Evgeny Lifshitz, and others, and has become standard. The use of non-covariant objects like pseudotensors was heavily criticized in 1917 by Erwin Schrödinger and others. Following his research on general relativity, Einstein entered into a series of attempts to generalize his geometric theory of gravitation to include electromagnetism as another aspect of a single entity. In 1950, he described his "unified field theory" in a Scientific American article entitled "On the Generalized Theory of Gravitation".[160] Although he continued to be lauded for his work, Einstein became increasingly isolated in his research, and his efforts were ultimately unsuccessful. In his pursuit of a unification of the fundamental forces, Einstein ignored some mainstream developments in physics, most notably the strong and weak nuclear forces, which were not well understood until many years after his death. Mainstream physics, in turn, largely ignored Einstein's approaches to unification. Einstein's dream of unifying other laws of physics with gravity motivates modern quests for a theory of everything and in particular string theory, where geometrical fields emerge in a unified quantum-mechanical setting. Einstein collaborated with others to produce a model of a wormhole. His motivation was to model elementary particles with charge as a solution of gravitational field equations, in line with the program outlined in the paper "Do Gravitational Fields play an Important Role in the Constitution of the Elementary Particles?". These solutions cut and pasted Schwarzschild black holes to make a bridge between two patches. If one end of a wormhole was positively charged, the other end would be negatively charged. These properties led Einstein to believe that pairs of particles and antiparticles could be described in this way. In order to incorporate spinning point particles into general relativity, the affine connection needed to be generalized to include an antisymmetric part, called the torsion. This modification was made by Einstein and Cartan in the 1920s. The theory of general relativity has a fundamental law—the Einstein equations which describe how space curves, the geodesic equation which describes how particles move may be derived from the Einstein equations. Since the equations of general relativity are non-linear, a lump of energy made out of pure gravitational fields, like a black hole, would move on a trajectory which is determined by the Einstein equations themselves, not by a new law. So Einstein proposed that the path of a singular solution, like a black hole, would be determined to be a geodesic from general relativity itself. This was established by Einstein, Infeld, and Hoffmann for pointlike objects without angular momentum, and by Roy Kerr for spinning objects. Einstein conducted other investigations that were unsuccessful and abandoned. These pertain to force, superconductivity, gravitational waves, and other research. In addition to longtime collaborators Leopold Infeld, Nathan Rosen, Peter Bergmann and others, Einstein also had some one-shot collaborations with various scientists. Einstein and De Haas demonstrated that magnetization is due to the motion of electrons, nowadays known to be the spin. In order to show this, they reversed the magnetization in an iron bar suspended on a torsion pendulum. They confirmed that this leads the bar to rotate, because the electron's angular momentum changes as the magnetization changes. This experiment needed to be sensitive, because the angular momentum associated with electrons is small, but it definitively established that electron motion of some kind is responsible for magnetization. Einstein suggested to Erwin Schrödinger that he might be able to reproduce the statistics of a Bose–Einstein gas by considering a box. Then to each possible quantum motion of a particle in a box associate an independent harmonic oscillator. Quantizing these oscillators, each level will have an integer occupation number, which will be the number of particles in it. This formulation is a form of second quantization, but it predates modern quantum mechanics. Erwin Schrödinger applied this to derive the thermodynamic properties of a semiclassical ideal gas. Schrödinger urged Einstein to add his name as co-author, although Einstein declined the invitation.[161] In 1926, Einstein and his former student Leó Szilárd co-invented (and in 1930, patented) the Einstein refrigerator. This absorption refrigerator was then revolutionary for having no moving parts and using only heat as an input.[162] On 11 November 1930, U.S. Patent 1,781,541 was awarded to Albert Einstein and Leó Szilárd for the refrigerator. Their invention was not immediately put into commercial production, and the most promising of their patents were acquired by the Swedish company Electrolux.[163] The Bohr–Einstein debates were a series of public disputes about quantum mechanics between Albert Einstein and Niels Bohr who were two of its founders. Their debates are remembered because of their importance to the philosophy of science.[164][165][166] In 1935, Einstein returned to the question of quantum mechanics. He considered how a measurement on one of two entangled particles would affect the other. He noted, along with his collaborators, that by performing different measurements on the distant particle, either of position or momentum, different properties of the entangled partner could be discovered without disturbing it in any way. He then used a hypothesis of local realism to conclude that the other particle had these properties already determined. The principle he proposed is that if it is possible to determine what the answer to a position or momentum measurement would be, without in any way disturbing the particle, then the particle actually has values of position or momentum. This principle distilled the essence of Einstein's objection to quantum mechanics. As a physical principle, it was shown to be incorrect when the Aspect experiment of 1982 confirmed Bell's theorem, which had been promulgated in 1964. While traveling, Einstein wrote daily to his wife Elsa and adopted stepdaughters Margot and Ilse. The letters were included in the papers bequeathed to The Hebrew University. Margot Einstein permitted the personal letters to be made available to the public, but requested that it not be done until twenty years after her death (she died in 1986[167]). Barbara Wolff, of The Hebrew University's Albert Einstein Archives, told the BBC that there are about 3,500 pages of private correspondence written between 1912 and 1955.[168] Corbis, successor to The Roger Richman Agency, licenses the use of his name and associated imagery, as agent for the university.[169] In the period before World War II, The New Yorker published a vignette in their "The Talk of the Town" feature saying that Einstein was so well known in America that he would be stopped on the street by people wanting him to explain "that theory". He finally figured out a way to handle the incessant inquiries. He told his inquirers "Pardon me, sorry! Always I am mistaken for Professor Einstein."[170] Einstein has been the subject of or inspiration for many novels, films, plays, and works of music.[171] He is a favorite model for depictions of mad scientists and absent-minded professors; his expressive face and distinctive hairstyle have been widely copied and exaggerated. Time magazine's Frederic Golden wrote that Einstein was "a cartoonist's dream come true".[172] Einstein received numerous awards and honors, including the Nobel Prize in Physics.

Appendix B4: CAWE Ranked Results (Albert Einstein)

Only the top 25 results are shown as the whole output is too large to fit into this document.

CAWE Rank	Word Similarity	Name Similarity	Combined Similarity	Original URL	Item Type	Origin	Level	Search Engine Rank	CSS Selector Tags
1	0.67088	0.27068	0.47078	http://people.famouswhy.com/albert_brooks	HTML	Bing	2		all
2	0.45727	0.11990	0.28858	http://wn.com/Mel_Brooks	HTML	Yahoo	2		p
3	0.28268	0.15963	0.22116	http://www.ask.com/web?o=102140&l=dir&qsrc=6&ad=dirN&q=Albert Einstein%2	HTML	Bing	2		all
4	0.23336	0.17263	0.20299	http://festivals.iloveindia.com/teachers-day/famous-teachers/albert-einstein.html	HTML	Bing	1	254	p
5	0.24435	0.15604	0.20019	http://search.yahoo.com/bin/search?p=Albert Einstein Physicist	HTML	Google	2		all
6	0.23913	0.14564	0.19238	http://www.celebrities-galore.com/celebrities/albert-einstein/home/	HTML	Bing	2		all
7	0.25456	0.11139	0.18297	http://www.whosdatedwho.com/sections/celebrities/first-name/Albert	HTML	Bing	2		all
8	0.20391	0.15652	0.18021	http://en.wikipedia.org/wiki/Albert Einstein	HTML	Google	2		all
9	0.20858	0.14730	0.17794	http://www.softschools.com/timelines/albert_einstein_timeline/16/	HTML	Bing	1	336	all
10	0.23336	0.12227	0.17781	http://www.worldcat.org/identities/lccn-n79-22889	HTML	Bing	2		p
11	0.19909	0.15394	0.17651	http://www.answers.com/article/1282317/who-was-little-albert	HTML	Yahoo	2		all
12	0.18858	0.16339	0.17599	http://www.december212012.com/press_room.htm	HTML	Bing	2		p
13	0.20913	0.14032	0.17473	http://www.celebrities-galore.com/celebrities/albert-einstein/home/	HTML	Bing	2		p
14	0.30038	0.04323	0.17181	http://www.whosdatedwho.com/sections/celebrities/first-name/Albert	HTML	Bing	2		p
15	0.21698	0.12207	0.16952	http://libris.kb.se/auth/184709	HTML	Bing	2		p
16	0.23071	0.10814	0.16943	http://www.factmonster.com/biography/var/alberteinstein.html	HTML	Yahoo	1	61	all
17	0.21864	0.11979	0.16922	https://www.reviewessays.com/Biographies/Albert-Einstein/2257.html	HTML	Bing	1	241	all
18	0.18056	0.15785	0.16921	http://www.december212012.com/press_room.htm	HTML	Bing	2		all
19	0.20946	0.12693	0.16819	http://alberteinsteinblog.org/a-universal-visionary/	HTML	Yahoo	2		p
20	0.22061	0.11438	0.16750	http://www.answers.com/Q/What_did_Albert_Einstein_do	HTML	Bing	2		all
21	0.21833	0.11528	0.16680	http://www.amazon.com/gp/search?index=toys&keywords=Albert Einstein and R	HTML	Bing	2		all
22	0.19598	0.13713	0.16656	http://www.slideshare.net/lalydiaa/albert-einstein-5680391	HTML	Bing	1	236	all
23	0.22020	0.11235	0.16628	http://famous-relationships.topsynergy.com/Albert_Einstein/	HTML	Bing	1	258	p
24	0.22018	0.11148	0.16583	http://www.answers.com/Q/Where_was_Albert_Einstein_from	HTML	Bing	2		all
25	0.11964	0.21178	0.16571	http://www.ducksters.com/movies.php	HTML	Google	2		p

Appendix C1: Query (Syed Rizwan Farook)

Syed Rizwan Farook
Environmental health officer
Born: 1987, Chicago, Illinois, United States
Died: December 2, 2015, San Bernardino County, California, United States
Spouse: Tashfeen Malik (m. ?–2015)
Parents: Rafia Farook, Syed Farook
Siblings: Saira Khan, Syed Raheel Farook

Appendix C2: CAWE Ranked Results (Syed Rizwan Farook)

Only the top 25 results are shown as the whole output is too large to fit into this document.

CAWE Rank	Word Similarity	Name Similarity	Combined Similarity	Original URL	Item Type	Origin	Level	Search Engine Rank	CSS Selector Tags
1	0.75957	0.46325	0.61141	http://newsfoxes.com/2015/12/breaking-california-shooter-syed-farook-received-28500-righ	HTML	Bing	2		p
2	0.82946	0.32375	0.57660	http://www.latimes.com/local/lanow/la-me-san-bernardino-shooter-endured-turbulent-home	HTML	Bing	2		p
3	0.80163	0.34655	0.57409	http://www.nydailynews.com/news/national/photo-shows-san-bernardino-shooters-entering	HTML	Bing	2		p
4	0.67201	0.46057	0.56629	http://www.newsfoxes.com/2015/12/rafia-farook-5-fast-facts-you-need-to-know-about-san-b	HTML	Bing	2		p
5	0.78054	0.34816	0.56435	http://news.investors.com/photopopup.aspx?id=784360	HTML	Google	1	252	p
6	0.80342	0.31435	0.55889	http://transparentcalifornia.com/salaries/search/?q=syed+farook	HTML	Google	1	11	all
7	0.77497	0.31604	0.54550	http://www.nydailynews.com/news/national/photo-shows-san-bernardino-shooters-entering	HTML	Bing	2		all
8	0.76701	0.32327	0.54514	http://www.talkvietnam.com/2015/12/leaving-behind-baby-and-bombs-couple-sows-panic-ir	HTML	Bing	2		p
9	0.89888	0.19062	0.54475	http://www.ourhealthtogether.com/dailyentertainmentnews/tag/syed-raheel-farook-la-sierra-	HTML	Bing	2		all
10	0.85290	0.23167	0.54229	http://www.latimes.com/local/lanow/la-me-san-bernardino-shooter-endured-turbulent-home	HTML	Bing	2		all
11	0.58059	0.49237	0.53648	http://newsfoxes.com/2015/12/5-fast-fact-about-enrique-marquez-syed-farooks-best-friend-	HTML	Bing	2		p
12	0.71246	0.33425	0.52335	http://www.talkvietnam.com/2015/12/leaving-behind-baby-and-bombs-couple-sows-panic-ir	HTML	Bing	2		all
13	0.48273	0.53936	0.51104	http://www.newsfoxes.com/2015/12/5-fast-fact-about-enrique-marquez-syed-farooks-best-f	HTML	Bing	1	633	p
14	0.78630	0.23355	0.50992	http://www.ndtv.com/topic/san-bernardino-police-chief-jarrod-burguan	HTML	Google	2		p
15	0.82863	0.18286	0.50575	http://www.latimes.com/local/california/la-me-sb-farook-family-20151211-story.html	HTML	Google	1	15	p
16	0.72491	0.28082	0.50286	http://www.inquisitr.com/2608110/california-shooting-syed-farook/	HTML	Yahoo	2		p
17	0.82368	0.17674	0.50021	http://www.dailymail.co.uk/news/article-3348741/FBI-investigating-San-Bernardino-gunmar	HTML	Yahoo	2		p
18	0.64961	0.35055	0.50008	http://newsfoxes.com/2015/12/breaking-california-shooter-syed-farook-received-28500-righ	HTML	Bing	2		all
19	0.73463	0.26434	0.49948	http://www.firstpost.com/tag/farook	HTML	Yahoo	2		all
20	0.85327	0.13715	0.49521	http://www.inquisitr.com/2607263/san-bernardino-shooting-everything-known-about-shoote	HTML	Yahoo	2		p
21	0.58036	0.40202	0.49119	http://www.newsfoxes.com/2015/12/5-fast-fact-about-enrique-marquez-syed-farooks-best-f	HTML	Bing	1	632	p
22	0.79803	0.18383	0.49093	http://www.google.com/search?q=Syed+Farook	HTML	Bing	2		all
23	0.55450	0.42640	0.49045	https://www.linkedin.com/in/syed-farook-26204092	HTML	Google	2		p
24	0.76732	0.21320	0.49026	https://www.linkedin.com/in/syed-farook-26204092	HTML	Google	2		all
25	0.48315	0.49237	0.48776	http://www.newsfoxes.com/2015/12/5-fast-fact-about-enrique-marquez-syed-farooks-best-f	HTML	Bing	2		p

Appendix C3: CAWE Name Mining Results (Syed Rizwan Farook)

As the output is too large to fit this document, only the top 60 ranked names will be included.

Rank	Name	Type	Avg Weighted NF
1	San Bernadino Female	ORGANIZATION	14.95655681
2	Jesu	PERSON	13.11484894
3	Gibran Malik	PERSON	12.71307329
4	AEST Facebook Twitter Google	ORGANIZATION	11.88773833
5	Daily Digest	ORGANIZATION	9.018647249
6	Darcie Loreno	PERSON	7.973571513
7	Clay County Regional Events Center	ORGANIZATION	7.535296052
8	San Bernardino County Sheriffs Department	ORGANIZATION	7.31578361
9	Jay Smith	PERSON	7.292638184
10	San Bernardino Shootings Press Conference	ORGANIZATION	6.993436126
11	Imran Farooq Murder Case & Altaf Hussain Kya Hai Andar	ORGANIZATION	6.901832932
12	ET Facebook Twitter Google	ORGANIZATION	6.854144548
13	Associated Press Image	ORGANIZATION	6.818988557
14	Gibran	PERSON	6.730450567
15	Robert Bruce Spencer	PERSON	6.691025583
16	Sally Allan Watch	PERSON	6.495874739
17	Earth Watch	ORGANIZATION	6.495874739
18	San Bernadino Shooters Family Lawyers	ORGANIZATION	6.357669206
19	Rasheed Masood	PERSON	6.251714446
20	Cheryl Dorsey	PERSON	6.219866481
21	Tatiana Farook	PERSON	6.16659075
22	Watch San Bernardino	LOCATION	5.97069257
23	Muneeb Farooq	PERSON	5.689804231
24	Balochistan Case	LOCATION	5.689804231
25	Region San Bernardino	ORGANIZATION	5.61370405
26	Shooters San Bernardino	ORGANIZATION	5.476585707
27	San Bernadino County Sheriffs Department	ORGANIZATION	5.43354469
28	San Bernadino San Bernardino	LOCATION	5.347454788
29	Quron	LOCATION	5.347454788
30	Philip Coleman	PERSON	5.264531384
31	Farook	PERSON	5.042995509
32	Phil Taylor-Jones	PERSON	5.03606168
33	Anzuoni Reuters	ORGANIZATION	4.984408656
34	American Pastor	ORGANIZATION	4.978578702
35	Tom Mesereau	PERSON	4.788345387
36	Jim Scuitto	PERSON	4.777600355
37	San Bernardino Mass Shooting Goes Live Across Broadcast	ORGANIZATION	4.682065267
38	NIV	ORGANIZATION	4.582297614
39	ROBERT SPENCER	PERSON	4.553660404
40	Jihad Watch San Bernardino	LOCATION	4.553660404
41	al-Baghdadi Malik	PERSON	4.478019427
42	Oregon Mass Shooting	ORGANIZATION	4.36800675
43	Jimmy Kimmel Live Donald Trump	PERSON	4.354053396
44	Jay Dyer	PERSON	4.341667924
45	California Malik	PERSON	4.291026531
46	Central Time in Paltalk Christian	ORGANIZATION	4.270472904
47	San Bernardino The White House	LOCATION	4.267353173
48	Breitbart Breitbart News	ORGANIZATION	4.267353173
49	Sayed Raheel Farook	PERSON	4.259629483
50	Naresh Bhardwaj	PERSON	4.141099759
51	Ashok Gaikwad	PERSON	4.141099759
52	Jinnah Ground	LOCATION	4.141099759
53	Union of India	ORGANIZATION	4.141099759
54	Shahzaib Khanzada	PERSON	4.141099759
55	Justin Harris	PERSON	4.095086018
56	Megyn Kelly Full Segment	ORGANIZATION	4.095086018
57	Stacey Caster	PERSON	4.095086018
58	United States Reuters	ORGANIZATION	4.067778661
59	Lilly Ristic	ORGANIZATION	4.004500218
60	Baraghedeh	ORGANIZATION	3.939509552

Appendix C4: Results of Case Study 1 (Counter-Terrorism with Terrorist Profiling)

I) TERRORIST PROFILING

Personal Details (from Top 100 Relevant URLs)

Family Background

- Violent and fractured family with divorced parents
  - Rafia (mother) first filed divorce in 2006, stating her husband was abusive, alcoholic, and negligent, during which the children often had to intervene to "save [her]"
  - In Feb 2008, father threatened to kill himself during a fight with a son, and received a restraining order
  - Rafia later sought to dismiss divorce case in April 2008
  - Parents divorced in Oct 2008

Marriage and Future Family

- Met Tashfeen Malik (a Pakistani) on an online dating site
  - Another dating profile stated he was looking for one who took his Sunni Muslim faith seriously
- Travelled to Saudi Arabia for hajj, returned to California in July 2014 with Malik as his wife
  - He later started growing a beard, an early gesture of subscribing to jihad
- Had a daughter (6 months old at time of attack)
- Rafia moved in with them to be closer to family and to a mosque

Relevant Extracted Names Not in the Query:

Names	**Remarks**
Tatiana Farook	Wife of Syed Raheel Farook (Syed Rizwan Farook's brother)
Enrique Marquez	Long time best friend of Syed Rizwan Farook, whom with he planned - but did not carry out due to fear of being caught - a terror attack with in 2012. His two guns were amongst those used in the massacre.
Mariya Chernyk	Sister of Tatiana Farook, and wife of Enrique Marquez

II) COUNTER-TERRORISM

Possible Causes of Radicalisation

1. Violent upbringing, fractured family
2. Interaction with individuals supporting radical views
   - Farook contacted people who had been investigated for international terrorism, via social media and phone calls
   - Malik held extremist views before arriving at the US, and discussed jihad and martyrdom with Farook as early as 2013
   - Rafia (mother) is a member of a pro-caliphate Islamic group (Islamic Circle of North America) with links to a radical Pakistani political group (Jamaat-e-Islami)

Precursors before Attack

1. Unusual activity around his house
    - Received a number of packages in a short amount of time
    - Couple were doing a lot of work in their garage
    - Spotted by neighbour but not reported in fear of racial profiling
2. Support of radical religious views
    - Malik posted on Facebook that “we pledge allegiance”(presumably to ISIS) just before attack
    - Father told a newspaper (after the attack) that Farook agreed with the ideology of the leader of ISIS and was fixated against Israel
3. Transfers of suspicious amounts of money to and from his bank
    - $28,500 deposited into Farook’s bank account 2 weeks before attack
    - Withdrew $10,000 and deposited it into Union Bank branch in San Bernardino
    - Three $5,000 transfers to Rafia Farook’s bank account in days leading up to attack
4. Visitation of shooting ranges a few days before attack
    - Farook spent a few hours each session practising shooting with a shotgun and a rifle

Appendix D1: Query (Sabah Earthquake)

Source: http://news.asiaone.com/news/malaysia/sabah-tremors-five-climbers-mt-kinabalu-feared-dead

Friday, 5 June 2015 | MYT 2:59 AM

Sabah quake: Five climbers on Mt Kinabalu feared dead

KOTA KINABALU: Fears are growing that there might have been casualties in the aftermath of the earthquake that hit Sabah on Friday.

Officials, however, are not commenting on unconfirmed reports that at least five people on the summit area of Mount Kinabalu might have been hit by rockfalls after the mountain shook for nearly a minute.

They were said to be among about 100 climbers stranded on the summit.

Bomba and police teams, including dog squads, were seen heading towards the park area while helicopters were being sought to airlift stranded climbers and mountain guides.

Most of the climbers were descending after reaching the 4,095m peak before sunrise when the earthquake struck at 7.17am, with boulders and rocks from the granite surface dislodging and rolling down rapidly.

It is learnt that two chalets – Panar Laban and Laban Rata – were also damaged but officials are still trying to ascertain the severity of the damage.

Some reports claimed that the Panar Laban chalet had been destroyed.

Initial reports from the Malaysian Meteorological Services Department said the 5.9 magnitude quake struck 16km northwest of Ranau.

Appendix D2: CAWE Ranked Results (Sabah Earthquake)

CAWE Rank	Word Similarity	Name Similarity	Combined Similarity	Original URL	Item Type	Origin	Level	Search Engine Rank	CSS Selector Tags
1	0.75479	0.59216	0.67347	http://goo.gl/Qm3ZXY	HTML	Bing	2		p
2	0.76512	0.56854	0.66683	http://www.thestar.com.my/News/Nation/2015/06/05/Sabah-tremors-Mt-Kinabalu/	HTML	Google	2		p
3	0.55631	0.75000	0.65315	http://www.malaysiakini.com/news/300832	HTML	Google	2		all
4	0.48209	0.77696	0.62953	http://says.com/my/news/earthquake-in-sabah-and-kota-kinabalu	HTML	Google	1	17	p
5	0.33877	0.85475	0.59676	http://sglinks.com/pages/51362958-sabah-quake-singaporean-girl-killed-at-mount-kinabalu-says	HTML	Yahoo	2		all
6	0.39534	0.79259	0.59397	http://singaporeseen.stomp.com.sg/singaporeseen/this-urban-jungle/unsung-heroes-mountain-gu	HTML	Yahoo	2		p
7	0.29191	0.89113	0.59152	http://multimedia.asiaone.com/node/200321/popup?page=31&maxWidth=95%&maxHeight=55%	HTML	Bing	2		all
8	0.38043	0.79559	0.58801	http://wn.com/Earliest_Footage_Of_Sabah_Earthquake_On_Mount_Kinabalu	HTML	Bing	2		p
9	0.30380	0.86603	0.58491	http://dragonpotatoes.tumblr.com/post/120912870771/singaporean-primary-school-kids-among-s	HTML	Bing	2		p
10	0.31267	0.84795	0.58031	http://whotalking.com/blog/earthquake in sabah	HTML	Bing	2		all
11	0.36706	0.78978	0.57842	http://wn.com/Sabah_Quake_Mt_Kinabalu_Guides_Rush_To_Aid_Despite_Aftershocks	HTML	Bing	2		p
12	0.37606	0.77778	0.57692	http://hype.my/2015/63221/sabahquake-kota-kinabalu-parts-sabah-rattled-earthquake/#comment	HTML	Yahoo	2		p
13	0.36185	0.78534	0.57359	http://sglinks.com/pages/51323233-sabah-quake-doctor-confirms-at-least-one-death-on-mount	HTML	Yahoo	2		all
14	0.46082	0.67952	0.57017	http://www.climbingmtkinabalu.com/2015/06/earthquake-at-mount-kinabalu.html	HTML	Yahoo	1	151	all
15	0.34628	0.78335	0.56482	http://www.dailyexpress.com.my/news.cfm?NewsID=100622	HTML	Google	2		all
16	0.41468	0.71083	0.56276	https://weehingthong.wordpress.com/2015/06/05/sabah-earthquake-5-june-2015/	HTML	Yahoo	2		all
17	0.34845	0.77152	0.55998	http://greatermalaysia.com/tag/mount-kinabalu/	HTML	Google	2		p
18	0.22881	0.89113	0.55997	http://multimedia.asiaone.com/node/200321/popup?page=47&maxWidth=95%&maxHeight=55%	HTML	Bing	2		all
19	0.39557	0.72134	0.55845	http://tubead.net/watch/sabah-quake-mt-kinabalu-climbers-trickle-down-to-kinabalu-pa.S2WjecJh	HTML	Google	2		all
20	0.43787	0.67726	0.55757	http://wn.com/Earliest_Footage_Of_Sabah_Earthquake_On_Mount_Kinabalu	HTML	Bing	2		all
21	0.27341	0.83992	0.55667	https://vulcanpost.com/268491/sabah-16-earthquakes-in-20-years/	HTML	Bing	1	98	all
22	0.24254	0.86603	0.55428	http://multimedia.asiaone.com/node/200321/popup?page=0&maxWidth=95%&maxHeight=55%&	HTML	Bing	2		all
23	0.32191	0.78512	0.55351	http://lgmt.org/_db_backups/tag/sabah-earthquake-2015	HTML	Google	1	122	p
24	0.40939	0.69538	0.55238	http://www.channelnewsasia.com/news/asiapacific/climbing-activities-at-mt/1894676.html	HTML	Google	2		all
25	0.40939	0.69538	0.55238	http://www.channelnewsasia.com/news/asiapacific/emergency-holiday/1894676.html	HTML	Bing	2		all

Appendix D3: Timeline of Disaster (Sabah Earthquake)
Timeline Showing Death Toll and the Missing Over Time

Date	Time	Update	Death Toll
5 Jun 2015	7.17am	6.0 magnitude earthquake strikes Ranou, no initial damage reports issued	-
	3.00pm	Discovered 3 rest houses and hostels, Donkey's Ears monument damaged	-
	6.18pm	32 accounted for; 122 missing	-
	8.30pm	118 missing; 2 deaths (1 Malaysian, 1 Singaporean)	2
6 Jun 2015	8.43am	17 missing (6 Malaysians, 1 Chinese, 1 Japanese, 9 Singaporeans)	-
	1.48pm	9 deaths (unidentified)	11
7 Jun 2015	4.01pm	2 missing (2 Singaporeans); 5 deaths (unidentified)	16
10 Jun 2015	10.40am	2 deaths (2 Singaporeans)	18