

Khurana & Khurana, Advocates And Ip Attorneys

IFAIA Centre, S/20-22, Greater Noida Shopping Plaza, UPSIDC Site-IV, Kasna Road, Greater Noida - 201308, India

Phone: +91.120.2342010, Fax: 2342011 Website: www.khuranaandkhurana.com

Email: info@khuranaandkhuraana.com

Offices: Noida (New Delhi) . Pune (Mumbai)

# INVENTIVE DISCLOSURE - CONFIDENTIAL

1. **Proposed Title of the Invention**

Software Product for suggesting relevant questions for a query using a novel approach

1. **Proposed Abstract of the Invention**

Suggesting similar questions for a user query has many applications ranging from reducing search time of users on e-commerce websites, training of employees to holistic learning for students. Mainly two approaches are studied for finding the similarity namely syntactic and semantic. Syntactic similarity judges how many words are common among two questions and semantic similarity deals with whether the two questions logically mean the same or not. We propose a novel combined approach for determining textual similarity that introduces a robust weighted syntactic and semantic similarity index for determining similar questions for a particular query which the user has presented as an input to the system. An end to end product has been developed to accomplish the same. Comprehensive set of experiments have been carried out to justify the efficacy of the proposed approach over the existing literature. An easy to use User Interface has been developed for the proposed product. The product as a whole can be put into use for different purposes like - similar FAQs about the internal working of a company, suggestions of questions for e-commerce websites and suggestions of questions for a technical query.

1. **Key Words:**

Natural Language Processing, Information Retrieval, Lexical Similarity, Syntactic Similarity, Semantic Similarity, Natural Language Understanding.

1. **Background of the Invention:**

**What are the present technologies that exist in the field of your invention and what are the limitations of the same? (Present state of Art)**

Presently the open source softwares use libraries like Natural Language ToolKit which help to find text similarity. The state-of-the-art architecture uses:

1. Syntactic Similarity
2. Semantic Similarity
3. Knowledge Graph.

The limitations are as follows:

1. Syntactic system fails to suggest similar questions stated with synonyms or in other words questions that carry the same meaning.
2. Semantic system fails to distinguish between dissimilar questions, thereby exposing the user to content that is not intended or is irrelevant.
3. The knowledge graph does not use natural language communication and hence may fail to produce the best result.

The product proposed in this document uses a novel approach that combines syntactic and semantic techniques. By doing so, the limitations encountered in the individual approaches were removed.

1. **What problems does the invention address and how your Invention is able to overcome the limitations/ problems of the existing technologies?**

**Holistic learning** : Many times students while learning a topic or subject for the first time, are unable to comprehend the entire concept as a whole. Our invention is able to overcome this limitation in the following manner - when a student asks questions on an online portal, similar questions to their entered query will be suggested and would help in holistic development immensely.

**Crime Investigations** : During an investigation, the person questioning the suspect is asking questions on his/her knowledge of the crime. This may be limited to the person’s skill and may miss out on some crucial interrogation. Our invention can help the interrogator by suggesting questions similar to the one he has asked so that all the possible cases/scenarios are covered.

**E commerce Websites** : Today the importance of e commerce cannot be overstated, the main aim of such sites being revenue generation, if user id provided to questions similar to his/her it will lead to reduction in searching time and hence increase in profits.

**Company Training Portals** : The system implemented can help any new joiner by predicting the problems or queries that may arise in future for an employee and providing a solution beforehand proves to be very helpful.

1. **Detailed Explanation of the Invention along with working examples. Kindly provide an elaborated description of each and every aspect of the invention (product and/or process) in great detail.**

In today’s world almost every field is a witness to human and machine interaction, this interaction if done using human language i.e. english or any other finds its place under the domain of Natural Language Processing. The product invented aims at predicting the queries that any user may encounter and provides suggestions for the same, thereby increasing the ease of communication many folds. The product description is as follows:

The product receives a set of questions in an excel file with possible solutions on the inter/intra net depending on the expected use. When a user say an employee, customer, consumer or buyer interacts with the product to gain information regarding a topic, the product suggests relevant queries to the user. The product finds the similarity between the user query and the questions present in the database using a complex procedure that involves the combination of two approaches.

The product is two phases namely the learning phase and the usage phase, during the learning phase multiple excel files and queries are to be provided. Where an excel file has questions and ranks allocated to some or all of them with respect to a query by a domain expert. Using these sets i.e. query and corresponding ranked excel file the product learns the best suitable combination of syntactic and semantic similarity. This learning process is completely dynamic and helps to customize according to the user needs.

In the usage phase the product suggests similar questions for which intra/internet links can be provided in the excel file. The product provides these facilities :

1. Self Learning
2. Suggesting Questions
   1. Provide intranet link
   2. Dynamically search on web
3. Creating required graphs and logs

1. **Kindly attach drawings, reports, papers, charts or other materials that may aid in your description.**

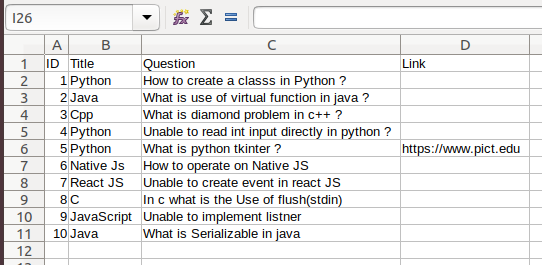
****

Image 1 (Database)

The image 1 shows a sample input file. The output shown in Image 2 contains clickable similar questions. On clicking a question user is directed to the link specified in the database, if the link field is kept empty then the system will generate a link to fire the query on a search engine.

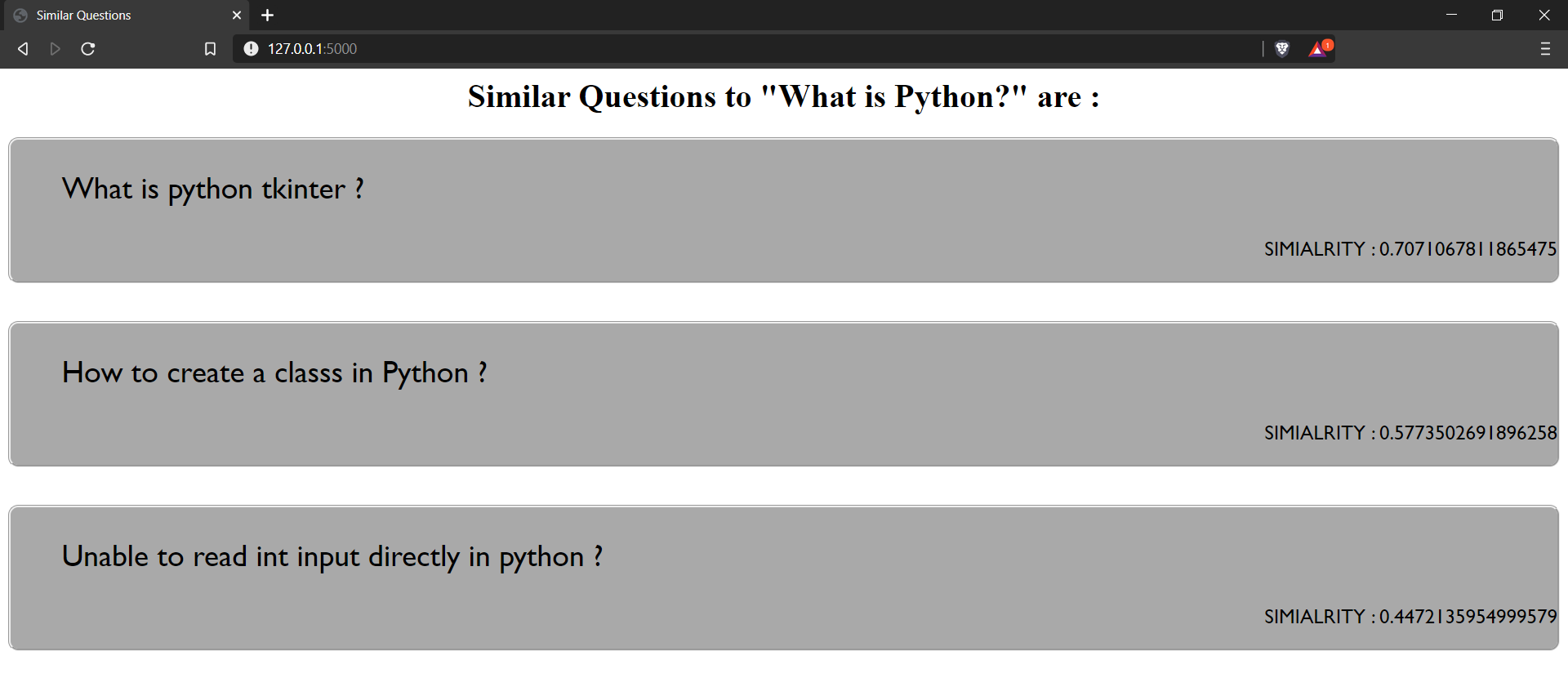


Image 2 (Output with User Interface)

Image 2 represents the output of the system along with the input query :

“ What is Python ? ”, for which the above similar questions are suggested.

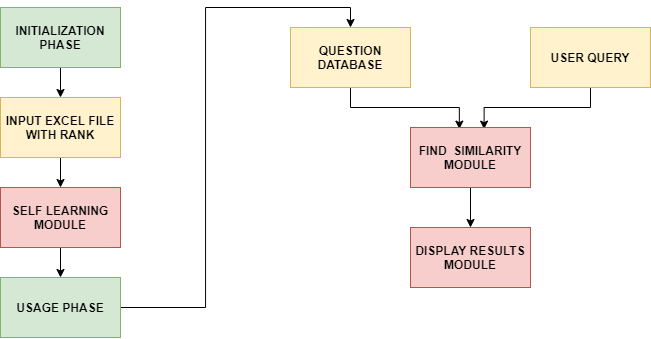


Image 3 (Product Overview)

Image 4 shows the proposed architecture which displays the methodology used to find the best value of lambda. Lambda indicates the combination of syntactic and semantic similarity score. The method shown in Image 4 is used in the self learning phase of the system. (In the image below, L represents Lambda, SSRD represents Sum of Squared Rank Distance)

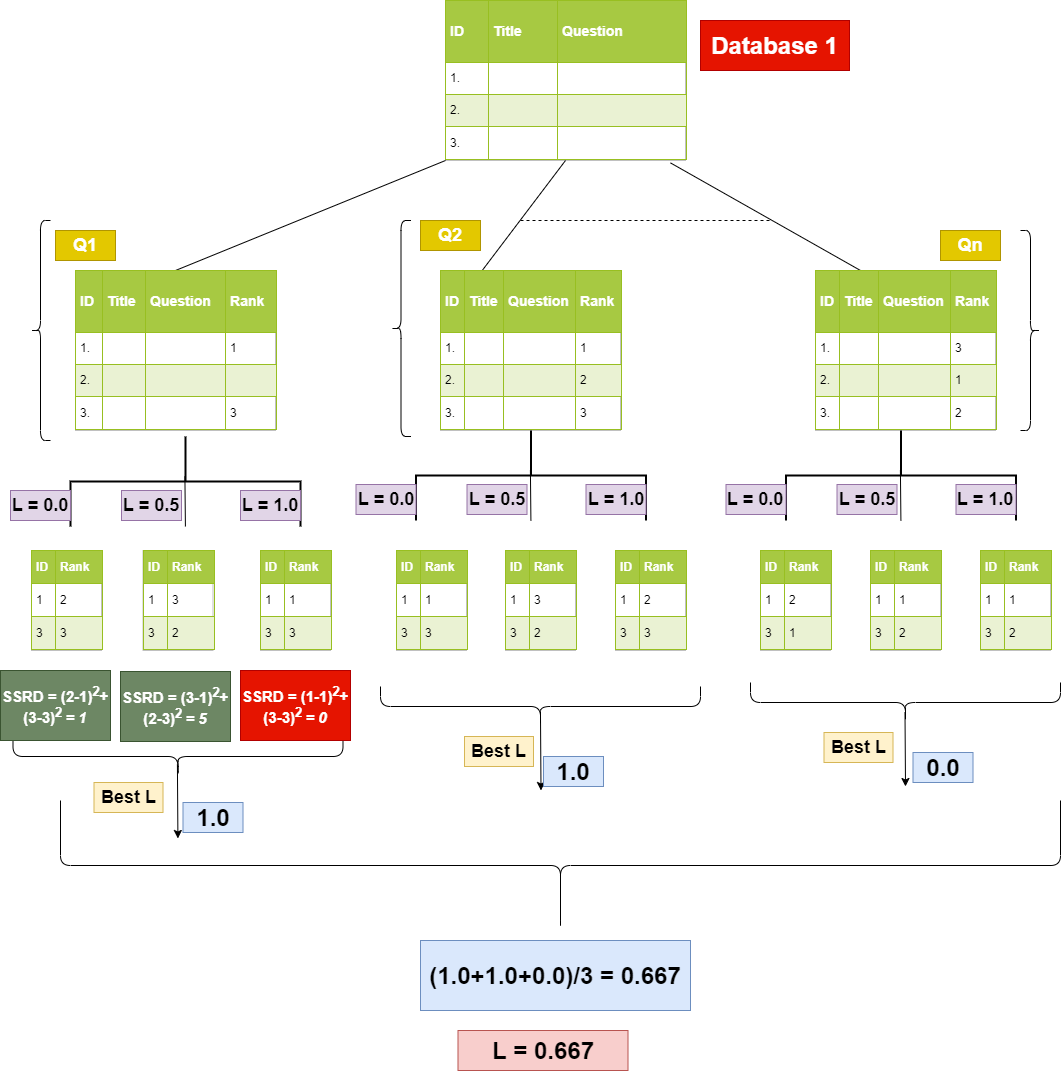


Image 4 (Self Learning Architecture)

1. **What are the aspects of your disclosure that you want to claim/monopolize?**

**Proposed Claims:**

1. Developed end to end product that uses a combined approach to suggest similar questions from the given database
2. Algorithm to dynamically find the optimal value of lambda for combined approach of similarity in order to suggest similar questions.
3. **Have you conducted novelty/inventiveness search for your invention? If yes, what are the databases /references used by you? What are the search results?**

Yes, we have conducted a novelty search for our invention. All the references used are mentioned below in the References section.

Article by Abdi et al(2015), proposes a combined approach for determining similarity.

We propose a product that automates the process suggested in the above mentioned article and also incorporates a new method of finding an optimum combination of the approaches. The proposed product generates the results without the need of human intervention, thus making the system user friendly and fast.

1. **Do you feel that a person of “average” skill (not-extraordinary skill) in your area of technology would have arrived at your invention with existing knowledge in public domain? If no, what could be the reasons for the same?**

We do not feel so. This invention needed skills which stood out from the mass skill set that the majority of Computer Science engineers possess. Having theoretical knowledge and practical skill will certainly help to an extent, but beyond that what is required is out-of-the-box thinking. We came up with this idea of combining two different approaches for suggesting questions into one and making it into a hybrid approach after months of deliberations. For a person of average skill, coming up with this idea is unlikely to come naturally to him/her.

This idea involved thinking about how a layman would use this product. It has a simple user interface to use and the selling point of the product is that it can be put to use in any field. This is the main reason we think a person with average skill would not be able to arrive at our invention.

1. **Kindly provide broad workable ranges for all the parameters involved in your invention.**

There are three parameters involved in the project namely the Number of questions in the input file, Number of questions to be displayed and Lambda.

Ranges of the Parameters:

1. Number of questions in the input file : 1 to (any number of the user’s choice).
2. Number of questions to be displayed : 1 to (Number of questions in input file).
3. Lambda : 0.0 to 1.0 ; where 0.0 indicates only use of semantic similarity and 1.0 indicates only use of syntactic similarity.
4. **References** **(if any)**
5. A. Abdi, N. Idris, R. M. Alguliev, and R. M. Aliguliyev, “Automatic summa-rization assessment through a combination of semantic and syntactic infor-mation for intelligent educational systems,”Information Processing & Man-agement, vol. 51, no. 4, pp. 340–358, 2015.
6. S. Bloehdorn and A. Moschitti, “Combined syntactic and semantic kernels for text classification,” in European Conference on Information Retrieval.Springer, 2007, pp. 307–318
7. W. H. Gomaa and A. A. Fahmy, “A survey of text similarity approaches,”International Journal of Computer Applications, vol. 68, no. 13, pp. 13–18,2013.
8. R. Mihalcea, C. Corley, C. Strapparavaet al., “Corpus-based and knowledge-based measures of text semantic similarity,” inAaai, vol. 6, no. 2006, 2006,pp. 775–780.
9. A. Islam and D. Inkpen, “Semantic text similarity using corpus-based word similarity and string similarity,”ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 2, no. 2, p. 10, 2008.
10. Medium, Introduction to Doc2Vec, 2017 (accessed January 20, 2020).[Online]. Available: <https://medium.com/wisio/a-gentle-introduction-to-doc2vec-db3e8c0cce5e>
11. StackOverflow,Word2Vec Time Complexity, 2019 (accessed March 3,2020). [Online]. Available:

<https://stackoverflow.com/questions/54950481/word2vec-time-complexity>

1. **Inventors Details (Full Names, Nationality and Addresses)**

|  |  |  |
| --- | --- | --- |
| Full Names | Nationality | Addresses |
| Shriniwas Nayak | Indian | Plot no. 319, Flat no. 2, Ramkrishna Towers, Laxmi Nagar, Nagpur - 440022 |
| Hrushabh Hirudkar | Indian | UG1 Bhaskar Arvind Apartment, Narendra Nagar, Nagpur- 440015 |
| Anuj Kanetkar | Indian | 44-A, Hill Road, Gokulpeth, Nagpur - 440010 |
| Archana Ghotkar | Indian | #TODO |

1. **Applicant Details (Full Names, Nationality and Addresses)**

|  |  |  |
| --- | --- | --- |
| Full Name | Nationality | Addresses |
| Archana Ghotkar | Indian | #TODO |

1. **Any additional notes or remarks.**

For any clarification regarding the content of the article please feel free to contact on the given details:

1. Shriniwas Nayak, email : [shriniwasnayak1@gmail.com](mailto:shriniwasnayak1@gmail.com), Mob no. : 9423531629
2. Hrushabh Hirudkar, email : [hrushabh.hirudkar@gmail.com](mailto:hrushabh.hirudkar@gmail.com), Mob no.:8805023410
3. Anuj Kanetkar, email : [anujkanetkar@gmail.com](mailto:anujkanetkar@gmail.com), Mob no. : 9403215199
4. Archana Ghotkar, email :