

# Ranking Model with a Reduced Feature Set for an Automated Question Generation System

**Manisha Satish Divate**

Department of Computer science  
University of Mumbai  
India  
divate.manisha.79@gmail.com

**Ambuja Salgaonkar**

Department of Computer science  
University of Mumbai  
India  
ambujas@gmail.com

## Abstract

A metric has been proposed to automatically generate well-formedness rank for machine generated questions. The grammatical correctness of a question, challenge due to the negation in the source text and number of transformations required to convert an ill-formed (not appealing to humans) question to a well-formed (meaningful and human appealing) question seem the prominent features.

We used 135 questions generated by Heilman and Smith's system (Heilman and Smith, 2011), developed a regression model using our feature set and tested it. The R-squared value is 93.5% which is quite acceptable. Relevance of the model has been corroborated by means of the residual plot. (Heilman and Smith, 2011) Metric is of 187 features and that of (Liu, 2012) is of 11. We suggested 16 features for a general automatic question quality enhancer to look into. However, the present model has been built by considering only 10 out of them. (Heilman and Smith, 2011) generated questions do not possess any infirmities indicated by the remaining ones. 75 questions were used in training and 60 were tested. Recognition accuracy is 94.4%.

## 1. Introduction

Automatic question generation systems (AQG) have been attempted since the mid-1970s (Wolfe J.H., 1976). A significant amount of work has been reported during the last decade, see details in Sec-

tion 2 below. To a certain extent AQGs have assisted human question-paper setters, but the quality of the automatically generated questions is often found to be inadequate. The strength and weaknesses of AQGs and a few remedies to enhance the quality of the questions have been discussed by us in (Salgaonkar and Divate, 2013).

It is necessary to formally compute the worth of an AQG by employing a quality metric. Developing such a metric for a given set of questions and therefore for an AQG is a research problem. The quality metric would serve the purpose of evaluating any automatic system that claims to provide quality enhancement of a set of questions. Call such a system an "automatic question quality enhancer" (AQQE). In the present paper we propose a new quality metric for questions generated by an AQG, suggest a framework for an AQQE, and demonstrate the improvement in quality due to this framework.

It goes without saying that the attributes of a question are the arguments of the quality function, and that, to improve the quality of a question, the attribute values have to be changed appropriately. Therefore the task of the AQQE is to rearticulate a given question in such a way that the changed attribute values yield a higher quality according to the metric. To improve the productivity of an AQG, our aim has to be optimization of the "objective function" of the metric, since we aspire to make a system to generate more and more questions that are acceptable to humans.

## 2. Previous Research

AQGs that attempt to measure *factual* as well as *deep learning*, including reading comprehension, writing ability and vocabulary, have been proposed in (Kunichika *et al.*, 2004; Brown *et al.*, 2005; Liu *et al.*, 2005; Chen *et al.*, 2006; Liu *et al.*, 2012). Systems generating MCQ type, fill-in-the-blank type or wh-type questions are another thrust area (Aist, 2001; Hashino and Hiroshi, 2005; Aldabe *et al.*, 2006; Mitkov, 2006; Prasad *et al.*, 2008; W. Chen *et al.*, 2009; Chen *et al.*, 2009; Heilman and Smith, 2009; Ali *et al.*, 2010; Liske, 2011; Agarwal and Mannem, 2011a; Agarwal and Mannem, 2011b).

Incorrect articulation of linguistic structures and an undue focus on relatively less important points are two commonly observed flaws in automatically generated questions. The statistics of the other types of errors observed in our experiments is given in the Table 1 below

Error Type	%
No error	43.10
Wrong WH	30.17
Vague	11.21
Formatting error	6.90
Ungrammatical	5.17
Does not make sense	4.31
Co reference not resolve	3.45

Table 1: Flaws observed in system generated questions

According to (Heilman and Smith, 2009), only 27.3% of the questions among the ones generated through their system are acceptable [Henceforth, the two authors are abbreviated as H&S]. The rank of a question has been computed by employing the linear regression model with 187 features (H&S, 2010). Among them, 53 features are as follows: length-based (3), language model (6), grammatical aspects (23), transformations (8), vagueness (3), negation (1), and wh-words (9). The remaining 134 are values computed from binary histograms of length and count features that serve as thresholds for controlling non-linearity among the numerical parameters used for describing syntax, length and linguistic features.

Empirical testing of the automatically generated questions revealed that out of the top ranked 20% (86 out of 428 questions), 52.3% were ac-

ceptable, where top rank is defined to be rank 3.5 and above on a 5-point scale, as selected by a team of human raters.

(Liu, 2012) is another notable work in the area of AQGs. This system helps in review writing and extracting citations from a given text. Further, the system categorizes the text and identifies the keywords that are used to fill in suitable question templates from the authors' question template library.

Among the approaches used in (Liu, 2012) to rank a question, two give the best results: point-wise logistic regression and pair wise support vector machines (the later technique is named RankSVM). With 11 features, when considering the top 25% ranked questions the acceptability of RankSVM was found to be 75.8%, while that of logistic regression was 74.2%, whereas 71.3 % of the top 50% are found acceptable. This figure is comparable to the result of (H&S, 2009). Also, it confirms the later result (H&S, 2011) that the increase in the percentage of acceptable questions is statistically significant when considering only the top ranked ones.

The 187 features of (H&S, 2010 and 2011) is arguably a large number, because a dozen features of (Liu, 2012) has led to a comparable result. Therefore the hypothesis arises whether it is possible to independently develop yet another reliable quality metric for questions using a model with a small number of features.

The findings about our 10 features regression model are reported in this paper. Section 3 is on nomenclature that is to set up a background for a novice reader. Details of our experiment are listed in Section 4. Observations and interpretations are presented in Section 5. Findings are summarized in the end of the paper.

### 3. Nomenclature

A question is *ill-formed*, i.e., unacceptable, if one or more flaws of the following types are observed in its articulation: incorrect grammar, semantic inadequacy, vagueness with respect to the answer, inadequate data, a wrong choice of WH form while presenting a question, or some editing is needed (H&S, 2009; Mannem *et al.*, 2010).

A question is *well-formed*, hence acceptable, if it is not ill-formed, i.e., it does not have any of the 6 flaws mentioned above.

A *Base sentence* is the source text for generating the questions.

An *answer phrase* is a machine-selected portion of the base sentence that is expected to be an answer to one or many potential questions that would be generated by an AQG.

Example (base sentence): *As is the case in a Parliamentary system, the government is formed by the party, alliance or group of assembly members who command the majority.*

(Answer phrase) *by the party, alliance or group of assembly members who command the majority*

IFQ (generated by HSAQG): *What is the government formed by?*

WFQ (manually generated): *Who forms the government in a Parliamentary system?*

Flaw: Wrong choice of wh-form.

#### 4. Methodology

Base text database consisted of five randomly selected paragraphs, of length 1 to 6 sentences, from wiki articles on a variety of topics: University of Mumbai, Lokmanya Tilak as a journalist, Government of Maharashtra, renewable energy, solar technology [Appendix 1].

135 questions were generated with HSAQG, out of which we randomly selected 75 as a training set. 13 questions (17.33% or a little over 1/6) were found to be redundant. Of the remaining 62, we found that 26 questions (42%) were acceptable to humans or WFQ, and 36 questions (58%) needed reframing in order to make them WFQs.

It appears that there is enough scope for further improvement in systems like HSAQG, hence this is a substantive research problem. Inspection of the data reveals that the changes required to improve the acceptability of the AQG-generated questions include removal of formatting errors, extra precision of questions to enhance clarity, more complete selection of answer phrases.

In the first step we analyzed the system generated IFQs with respect to the corresponding WFQs that were provided by a human expert. This exercise laid the foundation for generating the 16 di-

mensional vector representation of a question which is the core of our quality metric.

In the Table 2 below we list a few sample IFQs from our experiment, the corresponding WFQs and the flaws observed in each case.

Sentence simplification is another challenge in AQG systems. AQG system frames all possible questions from the modifier phrase, subordinate clause, appositives phrases, and from leading prepositional phrases. Those questions are unacceptable because they are *vague* or *grammatically incorrect*, or *do not make sense*. Our analysis reveals that AQGs fail to produce WFQs from the negative sentences.

In the next step we had a team of 5 human raters rank each of the questions on a 5-point scale. The protocol for generating the rank of a question was that the majority carries the vote. In case of much divergence in perceived rank, we took cognizance of others' views before choosing a rank. The sentiments of the raters were shared, which gave us an idea about what is not appealing to them. We formalized this feedback using the grammatical framework of English sentence structure. This exercise led to 16 rules for computing the rank of a question. These rules have been listed as R1 to R16. Only 10 rules (\*-marked) were applicable for the H&S AQG that we used. Examples for remaining 6 rules (which are not \*-marked) are taken from various internet sources. Ill-articulations are underlined and corrections are in bold.

R1(verbTenseInQuestion): Verb exists AND Verb Tense is correct => 1 else 0

Base sentence: *The Martians landed near the aqueduct.*

IFQ: *Where did the Martians landed?*

WFQ: *Where did the Martians **land**?*

\* R2(AuxVerb): AuxiliaryVerb is present in the base sentence AND (incorrect or no use of AuxiliaryVerb in the question) => 0 else 1

*The University of Mumbai was established 1857 by Dr John Wilson (after whom Wilson College in Mumbai is named), according to*

Sr No	Base Sentence	Answer phrase	IFQ	WFQ	Flaw	Flaw type
1	Maharashtra has bicameral legislature, i.e., it consists of two houses, Vidhan Sabha and Vidhan Parishad.	a bicameral legislature i. e.	What does Maharashtra consist of two houses-- Vidhan Sabha and Vidhan Parishad Maharashtra has?	Which type of legislature does Maharashtra have? Which legislature has two houses?	it refers to Maharashtra but, actually it should refer to the word bicameral	Inappropriate resolution of co-reference
2	Following this, on 22 June 1897, Commissioner Rand and another British officer, Lt. Ayerst, were shot and killed by the Chapekar brothers and their other associates	Commissioner Rand and another British officer	Who were shot? Who were killed?	Who were shot and killed A clearer question could be: Who were shot and killed on 22 June 1897 by the Chapekar brothers and their other associates?	Shoot and kill refers to an instance; Shoot and kill are not to be treated as two different instances.	Verbs come with a conjunction; question is too short
3	Tilak was born in a Chitpavan Brahmin family in Ratnagiri Maharashtra on 23 July 1856	23rd July 1856	When was Tilak born in a Chitpavan Brahmin family on 23 July 1856?	When was Tilak born ?	Date, family and place need to be treated as three different things	Additional context is to be removed
4	Tilak was born in a Chitpavan Brahmin family in Ratnagiri Maharashtra on 23 July 1856	on 23	What was Tilak born in a Chitpavan Brahmin family in Ratnagiri Maharashtra on July 1856?	When was Tilak born?	23 July 1856 is a phrase	Answer phrase is partially selected
5	Based on REN21's 2014 report, renewables contributed 19 percent to our energy consumption and 22 percent to our electricity generation in 2012 and 2013, respectively	to our energy consumption and 22 percent-	What did renewables contribute 19 percent to to our electricity generation in 2012 and 2013, respectively?	what [according to REN21's report] in 2012 renewables contributed 19 percent to?	Words order is incorrect. Question becomes ungrammatical.	Grammatically incorrect

Table Continued.....

Sr No	Base Sentence	Answer phrase	IFQ	WFQ	Flaw	Flaw type
6	Following this, on 22 June 1897, Commissioner Rand and another British officer, Lt. Ayerst were shot and killed by the Chapekar brothers and their other associates	yes/no	Did Commissioner Rand and another British officer follow this?	Did Chapekar brothers and their other associates kill and shoot Commissioner Rand and another British officer, Lt. Ayerst on 22 June 1897?	"Following this" indicates an event.	Does not make sense
7	Maharashtra has a bicameral legislature i. e. it consists of two houses-- Vidhan Sabha (legislative assembly) and Vidhan Parishad (legislative council).	yes/no	Does Maharashtra consist of two houses-- Vidhan Sabha and Vidhan Parishad Maharashtra has a bicameral legislature i. e.?	does maharashtra consist of two houses Vidhan sabha and Vidhan parishad?	Removal of "--" and "i.e" from question.	Formatting error
8	In 1792, Carey, a Baptist, who was not only a cobbler, but a linguist of the highest order.	Carey	Who was not only a cobbler, but a linguist of the highest order in 1792?	Who was Baptists, cobbler and linguist of highest order in 1792?	Frame affirmative question.	Negative sentence

Table 2: Sample IFQs with flaw types and suggested WFQ

*``Wood's dispatch'', drafted by Sir Charles*

IFQ : *Who was ``Wood's dispatch'' drafted by in 1854?*

WFQ: *Who **drafted** "Woods dispatch" ?*

\*R3(AnsPronoun): Correct Wh-phrase with respect to the Pronoun phrase in Answer => 1 else 0

Base Sentence: *It has 711 affiliated colleges*

IFQ: *What has 711 affiliated colleges?*

WFQ: *What has 711 affiliated colleges?*

\*R4(PronounAsSubjectInQuestion): Subject of the question phrase is pronoun => 0 else 1

Base Sentence: *It has 711 affiliated colleges*

IFQ: *What does it have?*

WFQ: *How many affiliated colleges does **University of Mumbai** have?*

\*R5(AnsPhraseAsPP): Answer phrase is a propositional phrase => 0 else 1

Base Sentence: *The Government of Maharashtra is the government for the state of Maharashtra in Western India.*

IFQ: *What is the Government of Maharashtra the government for the state of Maharashtra in?*

WFQ: *Where in India is the Government of Maharashtra the government for the state of Maharashtra?*

\*R6(SubordinateClause): Subordinate clause in question => 0 else 1

Base Sentence: *Solar technologies are broadly characterized as either passive solar or active solar depending on the way they capture, convert and distribute solar energy.*

IFQ: *What are Solar technologies broadly characterized as they capture, convert and distribute solar energy?*

WFQ: *What are the ways to characterize Solar Technologies?*

\*R7(NounPhraseInsidePP): Noun phrase that is the answer phrase is a part of a propositional phrase => 0 else 1

Base Sentence: *The Government of Maharashtra is the government for the state of Maharashtra in Western India.*

IFQ: *What is the Government of Maharashtra the government for in Western India?*

WFQ: *Which state in Western India the Government of Maharashtra is the government for?*

R8(SubjectMovement): Answer phrase is not a subject AND correct Subject-Auxiliary Verb movement => 1 else 0

Base Sentence: *Sam is reading the paper.*

IFQ: *what Sam is reading?*

WFQ: *What **is Sam** reading?*

R9(RemovedAppositives): Appositive phrase => 0 else 1

Base Sentence: *In June of 1987, The Bridge of Trinquetaille, Vincent van Gogh's view of an iron bridge over the Rhone, sold for \$20.2 million.*

IFQ: *what was Vincent van Gogh's view of an iron bridge over the Rhone sold for \$20.2 million?*

WFQ: *What was sold for \$20.2 million?*

R10(LeadConjunction): Leading conjunction phrase => 0 else

Base Sentence: *Since they had misbehaved, the boys were given one week suspensions from school.*

IFQ: *who had misbehaved since?*

WFQ: *who had misbehaved?*

R11(NounParticiple): Noun participle without the noun => 0 else 1

Base sentence: *Having been on the road for four days, the Todds were exhausted.*

IFQ: *Who were exhausted having been on the road for four days?*

WFQ: *Who were exhausted?*

\*R12(LeadingPP): Leading proposition phrase of a base sentence appearing in the question AND not moved to the end of the question => 0 else 1

Base sentence: *Based on REN21's 2014 report, renewables contributed 19 percent to our energy consumption and 22 percent to our electricity generation in 2012 and 2013, respectively*

IFQ: *Who were based on REN21's 2014 report?*

WFQ: *What [**according to REN21's 2014 report**] contributed 19 percent to our energy consumption in 2012?*

Base sentence: *For convocation, I will go to the University.*

IFQ: *For what will you go to the University?*

WFQ: *What will you go to the University for?*

WFQ: *Why will you go to the University?*

WFQ with a lesser rank: *When will go to the University? (The correct answer of this question will be "convocation day" or "on convocation day").*

\*R13(Wh-PhraseAsPerANS): Correct Wh-phrase with respect to the Answer in the Answer phrase => 1 else 0

Base Sentence: *Tilak was born in a Chitpavan Brahmin family in Ratnagiri, headquarters of the eponymous district of present day Maharashtra (then British India) on 23 July 1856.*

IFQ: *Who was headquarters of the eponymous district of present day?*

WFQ: ***what** was the headquarters of the eponymous district of present day Maharashtra?*

\*R14(LengthAnsPhrase): NOT (Answer phrase longer than 4 words OR Answer phrase lesser than 4) OR Answer phrase is a named entity => 1 else 0

Base Sentence: *The University of Mumbai offers Bachelors, Masters and Doctoral degrees apart from **diplomas and certificates in many disciplines**.*

IFQ: *What does the University of Mumbai offer Bachelors, Masters and Doctoral degrees apart from?*

WFQ: *Which courses does University of Mumbai offer along with Bachelors, Masters and Doctoral degrees?*

R15(ChkNegation): Base sentence is a negation AND Question form is negative => 0 else 1

If question is not affirmative then rank 0 else 1

Base sentence: *In 1792, Carey, a Baptist, who was not only a cobbler, but a linguist of the highest order.*

IFQ: *Who was not only a cobbler, but a linguist of the highest order in 1792?*

WFQ: *Who **was cobbler and linguist** of higher order in 1792?*

\*R16(Wh-PhraseChange): WFQ => 0 else number of modifications required to make it a WFQ

Base Sentence: *Maharashtra has a bicameral legislature i. e. it consists of two houses-- Vidhan Sabha (legislative assembly) and Vidhan Parishad (legislative council).*

IFQ: *Does Maharashtra consist of two houses-- Vidhan Sabha and Vidhan Parishad Maharashtra has a bicameral legislature i. e. ?*

WFQ: *does Maharashtra consist of two houses Vidhan sabha and Vidhan parishad?*

A question is represented as a 16 dimensional vector, of which all values but the last are either 0 or 1. Together with the human rated rank of each question as the dependent variable in the 17th place, we got a system of 62 equations in 16 independent variables (equivalently, 62 observations with 16 parameters). We attempted a solution of this system by employing a multiple linear regression.

## 5. Result and Discussion:

At 93.5%, the R-squared value, or the coefficient of determination, is satisfactory. We tested the model by using 61 observations that were independent of the training sample. Of which 6 questions were found redundant. The precision of the model from a sample of size 54 is 0.934.

As shown in the Figure 3 (a) and (b), the points in the residual plot when employed our model, are randomly dispersed. It confirms that the linear regression model is appropriate for this data.

We derived a quality metric for computing “how well-formed a question is”:

$$Y = 0.097 * AuxVerb + 0.194 * LeadingPP + 0.246 * LengthAnsPhrase + 0.017 * WhPhraseAsPerANS + 0.714 * AnsPronoun - 0.010 * PronounAsSubjectInQuestion - 0.215 * SubordinateClause + 0.208 * NounPhraseInsidePP - 0.037 * AnsPhraseAsPP - 0.588 * WhPhraseChange + 3.746$$

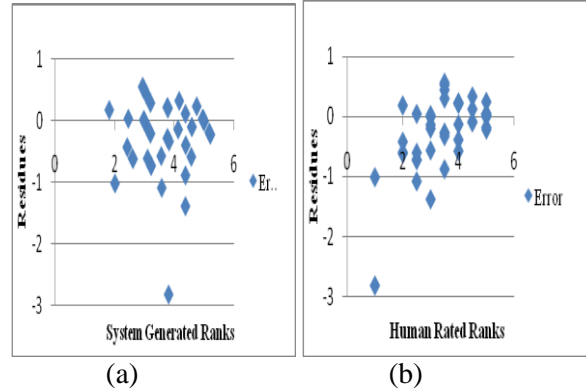


Figure 3 (a): Residual plot for ranks generated by LR model

Figure 3 (b): Residual plot for Human rated ranks

We observed a few interesting facts a majority of errors (30.17%) are because of wrong choice of Wh-phrase. It may be due to the fact that formulating a question using *what* is relatively easier and not incorrect per se. Therefore, AQGs employ this technique most of the time while generating a Wh-question. However, such questions are vague, imprecise so we classify them as IFQ. Articulating a question with a right Wh-phrase involves analysis of the answer term. Recall the example from [Section 4], namely, *Why will you go to the University?* The articulation of this WFQ requires the understanding that convocation is an event at the University and that has a purpose for its visitors. Interpretation of the event as a temporal entity gives rise to a less pertinent question: *When will you go to the University?*

In the case of an IFQ with a flaw, a wrong choice of Wh-phrase requires up to 5 modifications while transforming it into a WFQ. A list of sample templates for a variety of Wh-questions and the corresponding transformations is given below.

Sample question prototypes

Example1: <What Verb Object?>: *What drink chocolate of us?*

Example2: <What do-support Subject Verb?>: *What did you eat (today)?*

Example3: <How many Helping noun Verb Object?>: *How many of us drink chocolate?*

Example3: <How many Helping noun do-support Subject Verb?>: *How many modifications do we need?*

To change the Wh-phrase from a *what* [Example1] type question to *How many* [Example3] type we need to change the Wh-phrase and

the addition of a helping noun. We modify the question in 2 steps.

Ex: Base Sentence : *Tilak obtained his Bachelor of Arts in first class in Mathematics from Deccan College of Pune in 1877.*

Answer phrase: *in first class*

IFQ : *What did Tilak obtain his Bachelor of Arts in from Deccan College of Pune in 1877?*

WFQ: *which class did Tilak obtain in his Bachelor of Arts in Mathematics from Deccan College of Pune in 1877.*

To change IFQ to WFQ we need to perform changes like *What*→*Which* ; insert →*class* ; Add preposition → *in* before possessive noun.

We assign value 3 for 16<sup>th</sup> feature which indicates the number of modification required to make WFQ from given IFQ.

A too-long answer phrase leads to the formation of a wrong Wh-question, as it may contain sub-phrases or may lead to the formation of vague questions, since very little information remains in hand for formulating the questions. For example, with “*It is an elected government with 288 MLAs elected to the legislative assembly for a 5-year term*” as a base sentence and “*with 288 MLAs elected to the legislative assembly for a 5-year term*” as an answer phrase the choice of a question is “*Who is it an elected government with?*” lead to the formation of Wrong Wh-question.

The system generated ranks (R) labeled A to E (where A is the best, and E is not acceptable) as follows: A is 4 to 5, B is 3 to 4, C is 2 to 3, D is 1 to 2, E is 0 to 1 (upper bound is included, lower bound excluded). For classification we used J48 classifier, to classify the test data of 54 questions into defined five classes (A-E).

Systems performance is evaluated by calculating Precision and F1 measure using formula mentioned below.

$$\text{Precision (PPV)} = TP / (TP + FP)$$

$$F1 = 2TP / (2TP + FP + FN)$$

Study results shows average precision value for all classes is 0.934 and F1-score as 0.944.

Towards the conclusion of our experiment, we manually modified each IFQ such that its vector representation will yield an optimum value and computed its rank by applying H&S metric as well as our metric. There were 4, 13, 18 and 20 questions were of the ranks up to 2, 3, 4 and 5 respectively. It revealed that the ranks of all but the

question numbers 5 and 41 have been upgraded to 5. Rank of question number 5 and 41 were upgraded to 4. Answer phrase in these cases has been a pronoun phrase, *following this*. In the absence of background knowledge resolution of the pronoun reference is a challenge while framing a WFQ of rating 5.

## 6. Conclusion

This study reveals that researching AQGs is worth for their enhancement. HSAQG generated questions are free from the 6 types of infirmities, namely, correct verb tense, Subject-Auxiliary Verb movement, leading conjunction phrase, Appositive phrase, question form is negative. Our linear regression model with 16 features is simple and effective for ranking a question for its well-formedness. The inputs generated by it are worth looking while designing an AQQE. AQQE is a challenge that has immediate applications.

## References

- Agarwal, M., & Mannem, P. (2011a, June). *Automatic gap-fill question generation from text books*. In Proceedings of the 6th Workshop on Innovative Use of NLP for Building Educational Applications (pp. 56-64). Association for Computational Linguistics.
- Agarwal, M., Shah, R., & Mannem, P. (2011b, June). *Automatic question generation using discourse cues*. In Proceedings of the 6th Workshop on Innovative Use of NLP for Building Educational Applications (pp. 1-9). Association for Computational Linguistics.
- Aist G.(2001) *Towards automatic glossarization: automatically constructing and administering vocabulary assistance factoids and multiple-choice assessment*. International Journal of Artificial Intelligence in Education, 12, 212-231
- Aldabe, I., de Lacalle, M. L., Maritxalar, M., Martinez, E., & Uria, L. (2006, January). *Arikiturri: an automatic question generator based on corpora and nlp techniques*. In Intelligent Tutoring Systems (pp. 584-594). Springer Berlin Heidelberg.
- Ali, H., Chali, Y., & Hasan, S. A. (2010, June). *Automation of question generation from sentences*. In Proceedings of QG2010: The Third Workshop on Question Generation (pp. 58-67).



- Brown, J. C., Frishkoff, G. A., & Eskenazi, M. (2005, October). *Automatic question generation for vocabulary assessment*. In Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing (pp. 819-826). Association for Computational Linguistics.
- Chali, Y., & Hasan, S. A. (2012). *Towards Automatic Topical Question Generation*. In *COLING* (pp. 475-492).
- Chen, W. (2009). Aist, G., Mostow, J.: *Generating questions automatically from informational text*. In Proceedings of the 2nd Workshop on Question Generation (AIED 2009) (pp. 17-24).
- Chen, C. Y., Liou, H. C., & Chang, J. S. (2006, July). *Fast: an automatic generation system for grammar tests*. In Proceedings of the COLING/ACL on Interactive presentation sessions (pp. 1-4). Association for Computational Linguistics.
- Heilman, M., & Smith, N. A. (2009). *Question generation via overgenerating transformations and ranking*. (No. CMU-LTI-09-013). CARNEGIE-MELLON UNIV PITTSBURGH PA LANGUAGE TECHNOLOGIES INST.
- Heilman, M., & Smith, N. A. (2010b, June). *Good question! statistical ranking for question generation*. In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (pp. 609-617). Association for Computational Linguistics.
- Heilman, M., & Smith, N. A. (2010c, June). *Rating computer-generated questions with Mechanical Turk*. In Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk (pp. 35-40). Association for Computational Linguistics.
- Heilman, M., & Smith, N. A. (2009, July). *Ranking automatically generated questions as a shared task*. In The 2nd Workshop on Question Generation (Vol. 1, pp. 30-37).
- Heilman, M. (2011). *Automatic factual question generation from text* (Doctoral dissertation, Carnegie Mellon University).
- Hoshino, A., & Nakagawa, H. (2005, June). *A real-time multiple-choice question generation for language testing: a preliminary study*. In Proceedings of the second workshop on Building Educational Applications Using NLP (pp. 17-20). Association for Computational Linguistics.
- Kunichika, H., Katayama, T., Hirashima, T., & Takeuchi, A. (2004). *Automated question generation methods for intelligent English learning systems and its evaluation*. In Proc. of ICCE.
- Liske, H. (2011, June). *A framework for automated generation of examination questions from web based semantically treated search results*. In Proceedings of the 12th International Conference on Computer Systems and Technologies (pp. 510-515). ACM.
- Liu, M., Calvo, R. A., & Rus, V. (2010, January). *Automatic question generation for literature review writing support*. In Intelligent Tutoring Systems (pp. 45-54). Springer Berlin Heidelberg.
- Liu, M., & Calvo, R. A. (2011, January). *Question taxonomy and implications for automatic question generation*. In Artificial Intelligence in Education (pp. 504-506). Springer Berlin Heidelberg.
- Liu, M., Calvo, R., Aditomo, A., & Pizzato, L. A. (2012). *Using wikipedia and conceptual graph structures to generate questions for academic writing support*. Learning Technologies, IEEE Transactions on, 5(3), 251-263.
- Liu, M., Calvo, R. A., & Rus, V. (2012). *G-Asks: An intelligent automatic question generation system for academic writing support*. Dialogue and Discourse: Special Issue on Question Generation, 3(2), 101-124.
- Mannem, P., Prasad, R., & Joshi, A. (2010, June). *Question generation from paragraphs at UPenn: QGSTEC system description*. In Proceedings of QG2010: The Third Workshop on Question Generation (pp. 84-91).
- Mitkov, R., An Ha, L., & Karamanis, N. (2006). *A computer-aided environment for generating multiple-choice test items*. Natural Language Engineering, 12(02), 177-194.
- Pal, S., Mondal, T., Pakray, P., Das, D., & Bandyopadhyay, S. (2010). *Qgstec system description—juqgg: a rule based approach*. In Proceedings of the Third Workshop on Question Generation, pp.76-79
- Prasad, R., & Joshi, A. (2008, September). *A discourse-based approach to generating why-questions from texts*. In Proceedings of the Workshop on the Question Generation Shared Task and Evaluation Challenge (pp. 1-3).

Rus, V., & Lester, J. (2009, July). *The 2nd Workshop on Question Generation*. In AIED (p. 808).

Salgaonkar A. and Divate M.(2014c) *Ill-Formed to Well Formed Question Generator*. In ICCI'14, Mumbai.

Wolfe, J. H. (1976, February). *Automatic question generation from text-an aid to independent study*. In ACM SIGCUE Outlook (Vol. 10, No. SI, pp. 104-112). ACM.

## Appendix (source Wikipedia)

### Paragraph1:

The University of Mumbai offers Bachelors, Masters and Doctoral degrees apart from diplomas and certificates in many disciplines. The University of Mumbai is one of the largest universities in the world in terms of the number of students. In 2011 the total number of enrolled students was 5,49,432. It has 711 affiliated colleges. The University of Mumbai was established in 1857 by Dr John Wilson (after whom Wilson College in Mumbai is named), according to "Wood's dispatch", drafted by Sir Charles Wood in 1854. The name of the University has been changed from "University of Bombay" to "University of Mumbai" as per a gazette of the Government of Maharashtra dated September 4, 1996.

### Paragraph 2:

The Government of Maharashtra is the government for the state of Maharashtra in Western India. Maharashtra has a bicameral legislature i. e. it consists of two houses-- Vidhan Sabha (legislative assembly) and Vidhan Parishad (legislative council). It is an elected government with 288 MLAs elected to the legislative assembly for a 5-year term. The leader of the majority in assembly becomes the Chief Minister and selects the cabinet members. As is the case in a Parliamentary system, the government is formed by the party, alliance or group of assembly members who command the majority.

### Paragraph 3 :

Based on REN21's 2014 report, renewables contributed 19 percent to our global energy consumption and 22 percent to our electricity generation in 2012 and 2013, respectively. In 2011, the International Energy Agency said that "the development of affordable, inexhaustible and clean solar energy

technologies will have huge longer-term benefits. Renewable energy replaces conventional fuels in four distinct areas: electricity generation, hot water/heating, motor fuels, and rural energy services. Wind power is growing at the rate of 30% annually, with a worldwide installed capacity of 282,482 megawatts (MW) at the end of 2012, and is widely used in Europe, Asia, and the United States. Renewable energy is generally defined as energy that comes from resources which are naturally replenished on a human timescale such as sunlight, wind, rain, tides, waves.

### Paragraph4

Tilak was one of the first and strongest advocates of "Swaraj" (self-rule) and a strong radical in Indian consciousness. Tilak was born in a Chitpavan Brahmin family in Ratnagiri headquarters of the eponymous district of present-day Maharashtra (then British India) on 23 July 1856. He obtained his Bachelor of Arts in first class in Mathematics from Deccan College of Pune in 1877. In 1879 he obtained his LL.B degree from Government Law College. Following this, on 22 June 1897, Commissioner Rand and another British officer, Lt. Ayerst were shot and killed by the Chapekar brothers and their other associates.

### Paragraph 5:

Venus Express is outfitted with a camera that can take images in infrared and visual spectrums as well as a device called the Flux Gate Magnetometer, which searches for lightning signal

### Paragraph 6:

Solar technologies are broadly characterized as either passive solar or active solar depending on the way they capture, convert and distribute solar energy.