Query Expansion Evaluation for Chatbot Application

Laila Hidayatin
Information Technology Departement
State Polytechnic of Malang
Malang, Indonesia
lhidayatin@gmail.com

Faisal Rahutomo
Information Technology Departement
State Polytechnic of Malang
Malang, Indonesia
faisal@polinema.ac.id

***Abstract*—At this digital era, the recent stage of business system makes so many opportunity and innovation, such as online shopping. The system provides easier interaction between seller and customer. But the problem is, the seller can't response immediately while the customer asks. Therefore, a chatbot can be a solution so the seller can response a question quickly. This research proposes a chatbot system with a frequently asked question knowledge base. An additional query expansion mechanism based on thesaurus dictionary is implemented. Cosine similarity metric then being used to measure similary between query and question in frequently asked question. This research reveals the experimental results of this approach not as good as of similarity logic system.**

Keywords— chatbot, Indonesian, natural language processing, query expansion

I. INTRODUCTION

Recently, the development of business can make so many opportunities and innovations, such as online shopping. The system can make the interaction between seller and customer easier. But the problem still persists. The seller can not deliver their respond immediately while the customer sending some questions. Therefore, a chatbot system can be a solution to the seller, where they able to deliver a response to a question quickly.

Various techniques and approaches have been developed for Chatbot application [1][2]. But, evaluation papers of Indonesian language application is still limited. Therefore, this paper aims to explore further Indonesian chatbot evaluation to enrich the discourse of this topic. Previously, we have developed an automatic comment answering with natural language processing in Facebook. They use similarity calculation between customer comment and comment knowledge base with cosine similarity method. The accuracy is 46% [3]. The system has a low accuracy because it only can answer a comment with a same query from the knowledge base. Therefore, we add query expansion to the online store chatbot application with natural language processing. This method can expand terms searching with synonym searching, so it could improve the system answer performance even the question isn't same exactly with the knowledge base. Query expansion is a widely used technique in information retrieval system [4]. But the approach is not used yet for chatbot application in Indonesian.

To fill the open problem, this research evaluates the using of query expansion in Indonesian chatbot application. Hopefully, this research can produce a new program that more accurate and faster chatbot response. Besides that, it can minimize the risk of buy cancellation because of the slow response of services.

II. LITERATURE STUDY

A. Natural Language Processing

Natural Language Processing is a system that can identify a human-natural language with using of an artificial intelligence. It can convert human language to be understandable by a computer [3].

B. Query Expansion

Query expansion is a technique in information retrieval system to retrieve relevance feedback. The System adds the additional term query into the original query. Query expansion modifies a query to meet the information needed [4][5]. Query expansion technique can be divided into:

- Manual Query Expansion. A technique to modify the query by the user manually.
- Automatic Query Expansion. A technique to modify the query automatically. For example, a system that always adds new synonym from the first query can be classified as a automatic query expansion.
- Interactive Query Expansion. A technique that user has an interaction with the system in query expansion process.

C. Term Frequency *(TF)* – Inverse Document Frequency *(IDF)*

TF-IDF metric is used after stemming process to calculate the terms weight. This metric calculates *TF* and *IDF* for each term in a set of sentences. This metric calculates the *t* term weight in a *d* sentence with this formula [6]:

$$W_{dt} = TF_{dt} * IDF_t \qquad (1)$$

d : sentence id
t : term id
W : *d* sentences weight over the *t* term
TF : number of terms that searched in a sentence
IDF : Inversed Document Frequency

IDF value:

$$IDF = log \frac{D}{df} \qquad (2)$$

D : document total number in a corpus
df : document total number that contains searched term

D. Cosine Similarity

Cosine Similarity is a metric to calculate the similarity between **A** and **B** vector. The metric produces a cosine angle *x* between the two vectors. Cosine angle value between two vectors is determined as similarity value of two objects

compared. The minimum value is 0 (zero) and maximum value is 1 (one) [6]. Cosine similarity formula is defined below:

$$cos(\theta) = \frac{\boldsymbol{A}\cdot\boldsymbol{B}}{||\boldsymbol{A}||||\boldsymbol{B}||} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} {A_i}^2}\sqrt{\sum_{i=1}^{n} {B_i}^2}} \qquad (3)$$

With $\mathbf{A} \cdot \mathbf{B}$ as a dot product that represents the angle between of two vectors. The dot product is a scalar value as a result of calculation from two vectors that have the same component. If vector **A** and **B** has n-component, therefore dot product can be calculated with the formula below

$$\boldsymbol{A} \cdot \boldsymbol{B} = A_x * B_x + A_y * B_y + A_z * B_z \qquad (4)$$

$|\mathbf{A}|$ is a vector length. Vector length can be found by this formula:

$$|\boldsymbol{A}| = \sqrt{{X_1}^2 + {X_2}^2 + {X_3}^2} \qquad (5)$$

III. Query Expansion For Chatbot System

A. Data Gathering Method

Data gathering method is used to get the required data as a supporting component in the development of online store chatbot application. Some method that can be used to gather the data such as:

- Literature study to gather and collect references from many sources that have the required data.
- In this research, online store observation to gather data from customer's question from Funky Toys in online store website Tokopedia and Facebook fanpage @funkytoysgundam.

B. System Design

The system will implement a query expansion with NLP that illustrated in Fig 1. The customer sends a message to Telegram user @funkytoysgundam_bot. The message contains sentences in Bahasa and then the query expanded with Thesaurus Dictionary (Bahasa) depend on the synonyms to expand the query.

Next step is pre-processing. Firstly, will do a method called tokenizing, which will separate the text depend on terms that compose it. After the list of terms retrieved, then each letter on each term will be converted to lowercase, it called case-folding. Case-folding converts letter into lowercase for the letter a to z only, otherwise will be deleted and regarded as a delimiter. After that, it searched basic word from the separated term. This process called stemming.

Then the next step is calculating the weight of term relation with question's text. It did with calculating the value of *TF* and *IDF*. And the result is vector term **A**.

A knowledge base is a list of question and answer. Pre-processing and term relation weight calculation with calculates *TF* and *IDF* will be done in this domain. Thus, that calculation will produce vector term **B**. The last step is calculating the distance between vector **A** and **B**. So, the result is cosine angle x between that two vectors.

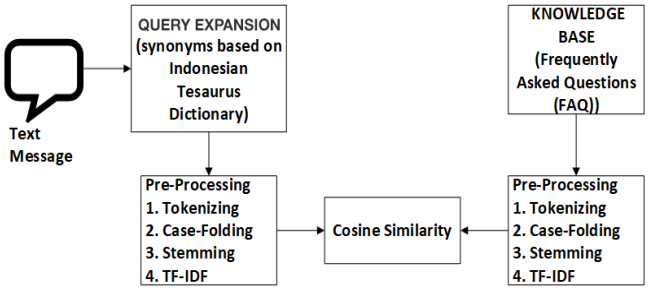

Fig 1. System overview

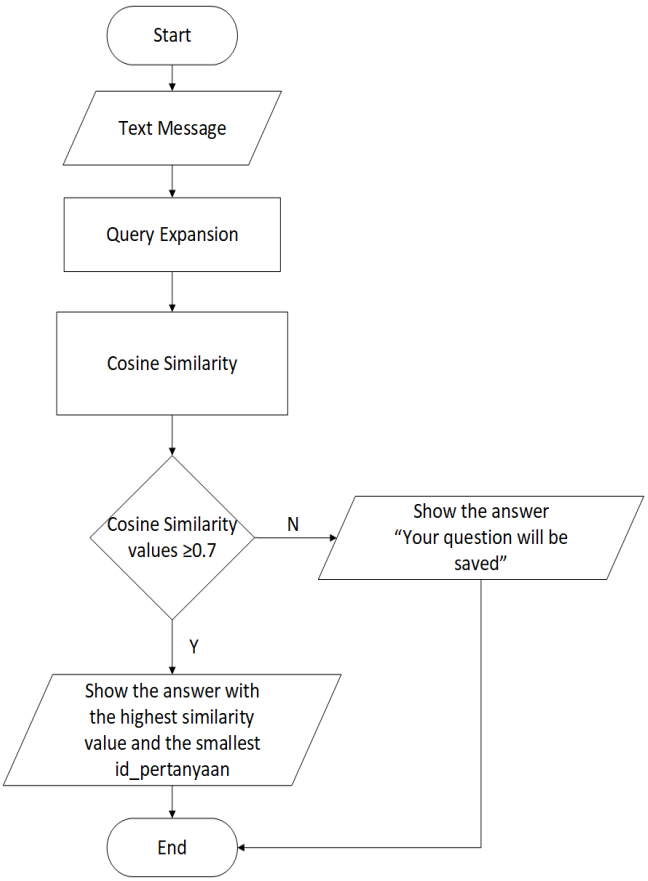

Fig 2. System flow design

C. Flowchart

Depends on the flowchart in Fig 2, the system's flow that customer can use chatbot application to ask about Funky Toys Gundam selling items with Telegram Messenger. The system will respond to the question with query expansion in Thesaurus Dictionary (Bahasa). After that, the system will compare the similarity between questions query and documents query with cosine similarity calculation.

If cosine similarity value is larger than or equal to 0.7, the system will respond back with an answer that matches with the customer's question. But if the cosine similarity value is smaller than 0.7, the system will respond back that the customer's question will be saved.

D. Implementation

This section explains the implementation of database, Telegram API platform, architecture system and chatbot Telegram messenger.

Depends on the database Fig 3, explained about the Physical Data Model as a query expansion object, storing data frequently asked questions and calculating the value of (TF) and (IDF)

Architecture system on Fig 4, explained about the representation of a system when the user sends a message via Telegram Messenger and then will be pushed message via webhook to the server to get an answer with the highest similarity value. The Server will response back and send a message via Telegram Messenger API.

Telegram offers two kinds of APIs for developers. The Bot API allows easily create programs that use Telegram messages for an interface. This research uses @botfather to retrieve the Token API from Telegram Bot.

From the implementation of chatbot in Telegram Messenger. The customer asks the question to Funky Toys Gundam Fig 5. The system will answer based on the customer's question if the cosine similarity value is larger or equal to 0.7. The system will answer with "Pertanyaan anda akan kami tampung terlebih dahulu" (This message will be saved first). The answer showed like that because of cosine similarity is smaller than 0.7.

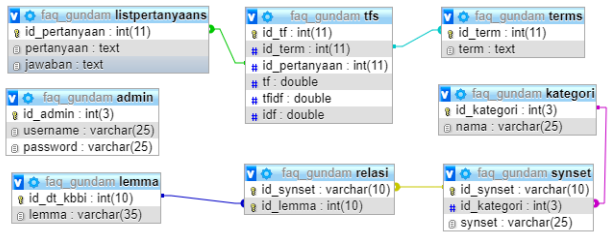

Fig 3. Physical data model

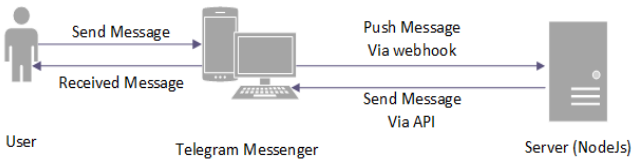

Fig 4. Architecture system

Fig 5. The bot answered the question

IV. Results And Discussion

In this result using three kinds of calculation, which is *Precision*, *Recall*, and $F-Score$ [7]. *Precision* is a level of performance between requested information and the answer. The *Recall* is a level of system performance in retrieving information. $F-Score$ is a combined value between *Precision* and *Recall*. The equations are described in equation (6), (7) and (8) respectively.

$$Precision = \frac{number\ of\ relevant\ document}{number\ of\ document} \quad (6)$$

$$Recall = \frac{number\ of\ relevant\ document}{number\ of\ document\ in\ collection} \quad (7)$$

$$F-Score = \frac{2(precision \times recall)}{(precision+recall)} \quad (8)$$

Chatbot application tested is with the comparison between implementation NLP with query expansion and without query expansion.

The result performance system with query expansion *Precision* of thresholds > 0.7, > 0.8 and > 0.9 are 0%, 0.48% and 3.33% respectively. *Recall* of thresholds > 0.7, > 0.8 and > 0.9 are 0%, 3.33% and 0.83% respectively. And $F-Score$ of thresholds > 0.7, > 0.8 and > 0.9 are 0%, 0.83% and 1.33% respectively. This result are described on Table 1.

The results performance system without query expansion *Precision* of thresholds > 0.7, > 0.8 and > 0.9 are 10%, 15% and 47.50% respectively. *Recall* of thresholds > 0.7, > 0.8 and > 0.9 are 6.11%, 11.67% and 25.67% respectively. And $F-Score$ of thresholds > 0.7, > 0.8 and > 0.9 are 7.22%, 12.29% and 30.80% respectively. This result are described on Table 2.

Depends on the testing result, overall the accuracy of chatbot application with NLP and query expansion is decreased.

Average *Precision* value is lower than without query expansion. In threshold > 0.7 decreased about 10%, threshold > 0.8 decreased about 14.52%, and threshold > 0.9 decreased about 44.17%.

Average *Recall* value is also lower than without query expansion. In threshold > 0.7 decreased about 6.11%, threshold > 0.8 decreased about 8.34%, and threshold > 0.9 decreased about 24.84%.

Average $F-Score$ value is lower than without query expansion. In threshold > 0.7 decreased about 7.22%, threshold > 0.8 decreased about 11.46%, and threshold > 0.9 decreased about 29.47%.

The results are not as good as our expectation because of several conditions. Firstly, the expansion of a query term is too wide. Because this system expands to all of the synonym set where the set includes the term. The system including so many terms that not really connected to the query term meaning. It can be limited by expanding the query more focus to the user meaning. Secondly, the system expands to all of the term types. The performance can be

meaning. It can be limited by expanding the query more focus to the user meaning. Secondly, the system expands to all of the term types. The performance can be increased further by limit the expansion to several types of the term, i.e. noun, verb, or adjective.

TABLE 1. TESTING RESULT WITH QUERY EXPANSION

Test	Threshold		
	≥0.7	**≥0.8**	**≥0.9**
$Precision$	0%	0.48%	3.33%
$Recall$	0%	3.33%	0.83%
$F-Score$	0%	0.83%	1.33%

TABLE 2. TESTING RESULT WITHOUT QUERY EXPANSION

Test	Threshold		
	≥0.7	**≥0.8**	**≥0.9**
$Precision$	10%	15%	47.50%
$Recall$	6.11%	11.67%	25.67%
$F-Score$	7.22%	12.29%	30.80%

V. CONCLUSIONS

Implementation of natural language processing with query expansion for online store chatbot application showed that query expansion with synonym searching in Thesaurus database gives us the result that the performance is decreased compared to the application without query expansion. Where

$Precision$ value with threshold $\geqslant$ 0.9 decreased about 44.17%, $Recall$ value with threshold $\geqslant$0.9 decreased about 24.84%, and $F-Score$ value with threshold $\geqslant$0.9 decreased about 29.47%.

This research can conclude that implementation of query expansion with cosine similarity has a low accuracy. We suggest for the next research is the improvement of Thesaurus dictionary database model. The threshold is also can be lowered and query expansion can be used to the specific category.

REFERENCES

[1] S. A. Abdul-Kader and J. Woods, "Survey on Chatbot Design Techniques in Speech Conversation Systems," *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 7, pp. 72–80, 2015.

[2] D. Aditya, S. Alisha, G. Darshana, D. Mrunmayi, and P. M. Joshi, "A Survey of Various Chatbot Implementation Techniques," *Int. J. Comput. Eng. Appl.*, vol. XI, no. Special Issue, 2017.

[3] F. A. Ramadhani, F. Rahutomo and P. Y. Saputra, "Rancang Bangun Pembuatan Sistem Aplikasi Penjawab Komentar Otomatis Facebook Menggunakan Metode NLP Sebagai Layanan Customer Service," in *Seminar Informatika Aplikatif (SIAP)*, Malang, 2017.

[4] H. K. Azad and A. Deepak, "Query Expansion Techniques for Information Retrieval: a Survey," *CoRR*, vol. abs/1708.00247, 2017.

[5] M. Ulum, F. Rahutomo and P. Y. Saputra, "Penerapan Query Expansion Dengan Database Tesaurus Pada Pencarian Koleksi Buku Perpustakaan Politeknik Negeri Malang," in *Seminar Informatika Aplikatif (SIAP)*, Malang, 2017.

[6] G. Salton, Automatic Text Processing, New York: Addison-Wesley, 1989.

[7] Y. Baeza and . N. Ribeiro, Modern Information retrieval, Harlow: Addison-Wesley, 1999.

[8] C. F. Pratama Putra., F. Rahutomo and P. Y. Saputra, "Implementasi Explicit Semantic Analysis Berbahasa Indonesia Menggunakan Corpus Wikipedia Indonesia," in *Seminar Informatika Aplikatif (SIAP)*, Malang, 2017.