

ICITDA 2020 - Upload Your Revised Paper

1 pesan

Google Formulir <forms-receipts-noreply@google.com>
Kepada: mulya@uii.ac.id

25 Oktober 2020 20.27

Terima kasih telah mengisi [ICITDA 2020 - Upload Your Revised Paper](#)

Inilah tanggapan Anda:

[Edit tanggapan](#)

ICITDA 2020 - Upload Your Revised Paper

Please revise your paper based on reviewers' comments, then prepare the following documents:

(1) Response letter that contains answers to comments/suggestions from reviewers. Please make it in the form of a table with two columns: reviewers' comments and your responses. Write down the changes you made (if any) and also show the location of the changes (chapter / sub-section / paragraph / picture / table number / etc.)

(2) Paper with the revised section is highlighted in yellow (.pdf)

(3) The latest version of the revised paper (.pdf)

Please upload document (3) to EasyChair.

Please upload docs (1) and (2) here.

We will review your paper revision.

The decision will be informed to you by email no later than five business days after you upload the revised paper.

If you have any question, feel free to ask us via: icitda@uii.ac.id

Thank you.

Best Regards,

Committee of ICITDA 2020

The International Conference on Information Technology and Digital Applications

Department of Informatics, Universitas Islam Indonesia

Web: <http://icitda.org>

Email: icitda@uii.ac.id

Alamat email *

mulya@uii.ac.id

Paper ID (in number) *

76

Author(s) name *

Sri Mulyati dan Kahfiara Krisna

Revised Paper *



I have updated my paper on EasyChair, according to the results of the last revision

Highlighted Paper *

Give a yellow highlight in the revised section, then save your document to .pdf. Maximum size: 10MB.

File yang dikirim



ICITDA_2020_paper_76 (3) - Sri Mulyati.pdf

Response Letter *

Response letter that contains answers to comments/suggestions from reviewers. Maximum size: 10MB.

File yang dikirim



responses - Sri Mulyati.pdf

We will review your documents. The decision of your submission will be sent to your email.
Thank you.

Buat Google Formulir sendiri



Sri Mulyati <135230506@uii.ac.id>

ICITDA 2020 submission 76 update

1 pesan

ICITDA 2020 <icitda2020@easychair.org>
Kepada: Sri Mulyati <mulya@uii.ac.id>

25 Oktober 2020 20.32

Dear authors,

we acknowledge that we received new files for your ICITDA 2020 submission. The information about this update is shown below.

Number: 76

Authors: Kahfiara Krisna and Sri Mulyati

Title: Document Searching of EPPS Test Result Using Indexing Method

Uploaded by: Sri Mulyati <mulya@uii.ac.id>

Updates:

paper, version 3 (1037946 bytes)

To access the new version of your submission you should log in to the ICITDA 2020 EasyChair page.



Sri Mulyati <135230506@uii.ac.id>

Fwd: ICITDA 2020 - Paper Acceptance Notification

3 pesan

Kahfiara Krisna <kkahfiara@gmail.com>
Kepada: Sri Mulyati <mulya@uii.ac.id>

28 Oktober 2020 11.32

Assalamualaikum wr wb.
Mohon maaf mengganggu bu. Saya ingin menyampaikan pesan dari icitda terkait registrasi pembayaran dan file upload.
Berikut pesan nya bu.

Terima kasih bu,
Wassalamualaikum wr wb

----- Forwarded message -----

From: **ICITDA 2020** <icitda2020@easychair.org>
Date: Mon, 26 Oct 2020 at 23:23
Subject: ICITDA 2020 - Paper Acceptance Notification
To: Kahfiara Krisna <kkahfiara@gmail.com>

Dear Mr./Ms. Kahfiara Krisna,

After we reviewed your revised paper, we would like to congratulate you on the acceptance of your paper!

Paper ID: 76
Title : Document Searching of EPPS Test Result Using Indexing Method

Thank you for your interest in The International Conference on Information Technology and Digital Applications (ICITDA) 2020.

PREPARING YOUR CAMERA-READY PAPER

Please ensure that you have followed the author guideline (including the references style) before preparing your camera-ready paper (described in <http://conferenceseries.iop.org/content/authors>), and submit the camera-ready paper as Microsoft Word or Latex file (and the corresponding compiled PDF) with the following format:

Format of file name: ICITDAcam-[Paper ID]-[First Author]-[Title]
Example: ICITDAcam-001-SRani-ThisIsTheTitleOfThePaper.pdf

REGISTRATION PAYMENT

All accepted papers are subject to registration fee as presenter (for more detail, please see: <https://icitda.org/information/registration/>).

Please make payment through bank transfer to the following account:

Bank Name: Bank Mandiri
Bank Address: Jl. Kaliurang km 14,5 Sleman, DIY 55584
SWIFT Code: BMRIIDJA
Account number: 137-00-1039369-8
Account name: Panitia Seminar FTI UII
Account Address: Jl. Kaliurang km 14,5 Sleman, DIY 55584

Or via Paypal:
taufiq.hidayat@uii.ac.id

Prepare your proof of payment with this format of file name: ICITDAslip-[Paper ID]-[Presenter]
Example: ICITDAslip-001-SRani.pdf

*Note: The paid registration fee is nonrefundable if you cannot participate in the conference.

VIRTUAL ATTENDANCE CONFIRMATION & FILES UPLOAD

Please submit your camera-ready paper, payment proof, and confirm your virtual attendance via <http://bit.ly/icitda-presenter> no later than November 2, 2020 (note that you need a Google/Google Apps account to access; please contact us if you do not have one or need assistance).

If you have any questions, feel free to reach us via icitda@uii.ac.id.
We look forward to meeting you virtually on November 13-14, 2020.

Best Regards,
Septia Rani
Program Committee of ICITDA 2020
The International Conference on Information Technology and Digital Applications
Department of Informatics, Universitas Islam Indonesia
<http://icitda.org>

Sri Mulyati, S.Kom.M.Kom <mulya@uii.ac.id>
Kepada: Kahfiara Krisna <kkahfiara@gmail.com>

28 Oktober 2020 13.34

siap dek, saya proses
[Kutipan teks disembunyikan]
--
Sri Mulyati

Kahfiara Krisna <kkahfiara@gmail.com>
Kepada: "Sri Mulyati, S.Kom.M.Kom" <mulya@uii.ac.id>

29 Oktober 2020 08.17

Baik bu, terima kasih.
Lalu untuk isi absen dan upload filenya, saya yang lakukan atau ibu ya?
[Kutipan teks disembunyikan]



Sri Mulyati <135230506@uii.ac.id>

ICITDA 2020 - Registrasi untuk Paper Internal Informatika UII

1 pesan

ICITDA UII <icitda@uii.ac.id>

31 Oktober 2020 11.04

Kepada: "Dhomas Hatta Fudholi, Ph.D." <hatta.fudholi@uii.ac.id>, Sheila Nurul Huda <sheila@uii.ac.id>, Septia Rani <septia.rani@uii.ac.id>, "Ahmad Fathan Hidayatullah, S.T." <fathan@uii.ac.id>, Rian Adam <195230101@uii.ac.id>, Ahmad Raf'ie Pratama <ahmad.raffie@uii.ac.id>, Sri Mulyati <135230506@uii.ac.id>, Ridho Rahmadi <ridho.rahmadi@uii.ac.id>, Fietyata Yudha <yudha@uii.ac.id>, Erika Ramadhani <erika@uii.ac.id>, Mukhammad Andri Setiawan <035230102@uii.ac.id>, Zainudin Zukhri <zainudin@uii.ac.id>, Irving V Papatungan <irving@uii.ac.id>

Salam Bapak Ibu,

Terkait dengan registrasi paper ICITDA 2020, ada beberapa hal yang ingin panitia sampaikan. Hal-hal tersebut sebagai berikut:

1. Untuk biaya registrasi paper yang diterima, dosen dapat mengajukan masing-masing pendanaannya melalui sim portofolio jurusan, dengan melampirkan bukti berupa email dari panitia ICITDA yang menyatakan bahwa revisi paper Bapak/Ibu sudah diterima (bukan yang masih pending revisions :D) dan file draft camera ready paper. Nominal biaya registrasi per paper adalah Rp 1.500.000,00 (meskipun paper tersebut nanti dipresentasikan oleh student Bapak/Ibu). Persetujuan atas pengajuan pendanaan ini ada di pihak jurusan.

2. Bapak/Ibu yang nanti akan mentransfer biaya registrasi ini ke rekening panitia ICITDA.

3. Di form registrasi presenter, isian untuk mengunggah bukti pembayaran kami buat bersifat mandatory (untuk menghindari dokumen dari para author yang tidak lengkap). Apabila pendanaan dari jurusan belum diterima, terdapat dua alternatif yang dapat Bapak/Ibu lakukan:

Alternatif 1: Bapak/Ibu melakukan pembayaran terlebih dahulu dengan biaya sendiri (menalangi dananya).

Alternatif 2: Bapak/Ibu mengunggah file pengganti yang berisi pernyataan bahwa pembayaran akan dilakukan menyusul.

4. Terkait camera ready paper, untuk keseragaman afiliasi publikasi, mohon gunakan konvensi berikut ini:

Dosen: afiliasi jurusan (Department of Informatics)

Mahasiswa: afiliasi prodi (Undergraduate/Master Program in Informatics)

Email mahasiswa gunakan: nama@students.uui.ac.id (jika memungkinkan).

Bagi Bapak/Ibu yang papernya masih tahap revisi (kemarin submit di batch final), instruksi detil mengenai proses registrasi akan dikirimkan saat sudah menyerahkan paper yang direvisi & revisinya sudah sesuai dengan yang diminta :D.

Mungkin itu dulu yang dapat kami sampaikan. Terima kasih atas perhatiannya.

Salam,

Septia Rani

Program Committee of ICITDA 2020

Document Searching of EPPS Test Result Using Indexing Method

Kahfiara Krisna

Department of Informatics
Islamic University of Indonesia
Yogyakarta

kkahfiara@gmail.com

Sri Mulyati

Department of Informatics
Islamic University of Indonesia
Yogyakarta

mulya@uii.ac.id

Abstract. This study focuses on documents searching on the EPPS test result, mainly administered in the education sector for selecting prospective students in universities' admissions process. There has been a considerable increase of admission data along with annually administered admission test for prospective students in every beginning of an academic year. Some faculties may administer the EPPS test in the second stage test of admission process after the prospective students pass the first stage in addition to having the general admission test. These several tests lead to a huge data accumulation with different information, making it difficult for the organizers to locate the document containing 15 aspects of required information in a comprehensive information display. Thus, it is necessary to apply an indexing method by having a predetermined index to retrieve required information in a particular document or to find other related documents by way of indexing method. This research used 74 pdf documents as the dataset. In particular, 60 datasets were used as references to retrieve phrases on the specified data, and 14 datasets were used to test data. The results revealed the most significant data with an average precision of 0.89, recall of 0.93 and f-measure of 0.91.

Keyword : indexing method, documents, searching

1. Introduction

The psychological test administrator has several documents stored in a computer directory. One of the documents stored is the EPPS test result. the number of test result documents is in accordance with the number of participants who completed the test. The EPPS (Edward Personal Preference Schedule) test is a psychological test that is performed by an individual test taker or by an organization. The EPPS test is widely used in the education sector, mainly to select prospective students in universities admission. Normally, the result of the EPPS test file will be stored as a pdf document containing the psychogram matrix and the concluding result of test taker performances.

As the new academic year begins, multitudes of prospective students apply for university admission based on their desired majors. The admission of countless applicants' data along with the result of their EPPS test will result in a huge pile of data that require large data storage in every test period. Given the large number of new prospective applicants' data and new information, it is difficult for the organizers to provide prospective students with accurate result of the selection process. The ineffective and incomprehensive data display requires the organizers to re-filter a number of documents and locate the particular search result according to the predetermined assessment. Moreover, this condition is worsened by absence of recorded data that can display new information according to the existing data in a more efficient presentation to be easily understood by both experts and organizers, which remains an obstacle at this time. At certain times, it is also difficult to locate certain documents containing required words. The only way to find certain information from a set of documents is by looking for the similar words contained in a document. This process is normally done manually, which is definitely time consuming, especially when there is a need to locate certain documents in large numbers.

Searching a certain document by way of an index method helps in finding indexes that contain the required aspect of EPPS test results and locating the index of the words searched for in the document. It is possible to classify the document based on its similarity through the method of word indexing from a set of documents. The indexing process aims to change an unstructured document into a structured document to ease the process of finding the index of the necessitated word and thus to display the document [1].

Given the above explanation, this research aims to propose a method to ease the organizers in locating required data and filter the number of potential applicants by finding the important and required information of the EPPS psychological test conclusions using the predetermined aspects of needs and by locating the position of the stored documents. This research was targeted for high school students majoring in Science who would apply for the Medical Faculty of UII.

On this account, the researchers tried to propose the indexing method as an efficient way for searching for information and document similarities to find suitable information regarding the required aspects contained in the EPPS test conclusions and to look for the similarity of recorded data with new data information in order to get the highest results based on the aspects needed by the organizer. This system, a dictionary or related meaning to aspects of need is developed in the measuring instrument used so that searches can be specific to the content of the document in the referenced directory. This research is formulated to answer the following research question, (1) how to find and display the information contained in the psychogram matrix results of the EPPS test to facilitate the search for the desired needs?

This paper is written based on the following structure of a research paper: the background, review of related works (literature), research methodology, the results and discussion, and the conclusions of this study.

2. Literature Review

Indexing method is the process of finding information from a large and unstructured document database in response to a required query [2]. The indexing process is carried out to display the desired information in the database by organizing the unorganized set of documents into a more structured document for an easy retrieval and easier documents finding based on the information contained in the required document [3].

This study addressed the EPPS (Edward Personal Preference Schedule) test as a form of personality test consisting of several answer choices to be selected by individuals according to the available questions. This test aims to find out personal needs that reflect the individual's personality [4]. This test was developed by Allen L. Edward by requiring us to select fifteen required aspects out of twenty compiled points that refer to Murray's personality concept. This personality test contains 225 questions, namely the compelling, objective and non-projective. The EPPS test is administered to observe the specific needs of an individual through 15 required aspects [5].

There have been several previous researchers on this topic as references of this study. The first research discussed the use of the indexing method in the search process for documents generated by manual search results. The advantage of index retrieval lies on structured data storage in order to store large amounts of various documents. The search process is done by creating a word index from the document [1].

The second study raised a topic on the search for text documents of archived letters using indexing and query methods. This research discussed the process of restoring information with an index so that documents can be retrieved when needed and can be located easily. It also used the index method to retrieve an index that has been registered in the database to find the desired sentence. With more detailed processing in the query development process, word grouping can be summarized more simply [2].

The third research entitled "Single Document Keywords Extraction in Indonesian using Phrase Chunking" [6]. This research discussed the formation of keywords from documents automatically using phrase concatenation. Phrase chunking was done using the Part of Speech (POS) pattern to extract phrases and keywords. Keywords were selected from the candidates based on the number of

words in the phrase and several types of phrase data forms. The shortest keywords with reduced strings were extracted from the abstract and sorted by their highest frequency values. This study conducted an experiment using types of phrases such as different sources of keywords, variations in phrase patterns and provided performance comparisons between types of phrases.

The fourth research discussed the use of a retrieval approach as a solution to existing problems by relying on concepts rather than keywords for indexing and retrieval. It aimed to retrieve documents that were relevant to a specific user request. Indexing and weighting of documents and queries were done using semantic entities, concepts, and keywords. In this approach, concepts were identified from the document and measured according to their frequency distribution and their similarity to other concepts in the document [7].

The technique used for text documents is widely used with the IR (Indexing and Retrieval) technique. The purpose of an IR system is to retrieve relevant material from a document database in response to a query [8]. The five research discussed the design of information retrieval system by employing Vector Space Model in the tracking and also similarity measures of Cosine Similarity as the method to rank the documents found that match with the keywords/query[9]

3. Research Metodology

We proposed an automatic keyword extraction technique using generate frase in Indonesian document. Figure 1 shows the methodology of this research.

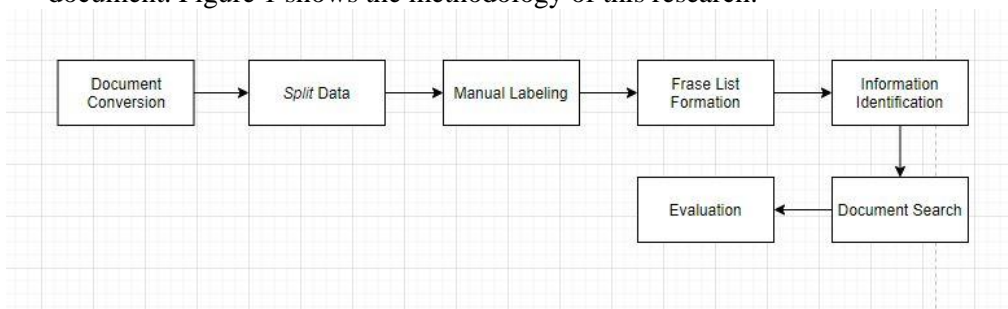


Figure 1. Research Metodology

3.1 Document Conversion

At this stage, the prospective applicants' psychological test results are collected into one EPPS folder in pdf form. The document will be an input data to be processed further. The EPPS test results contain two paragraphs of sentences relating to the applicant's most and least required aspects. The data on the EPPS test results will be taken by indexing the paragraph "description" that appears on the document and it will be processed further through the preprocessing process.

3.1.1 Preprocessing.

The preprocessing is an early stage process that is prepared to carry out further processing of the text in the data before going through the indexing process. The process in this preprocessing stage only includes several steps, namely casefolding (changing the front letter to lowercase) and tokenizing (separating each word into one line only). The results of the preprocessing will later be saved and converted to csv files. The CSV file is the result of the document conversion process and preprocessing of the EPPS test results.

3.2 Split Data

At this stage, the dataset is divided randomly into training data and test data with a ratio of 80%: 20%. The training data will be used as a source of formation of a list of phrases and labels, which will later be matched with the document. Meanwhile, test data is useful for looking for occurrences with a value similar to the recorded data.

3.3 Manual Labeling

At this point, the converted csv file will contain a table of "phrases" and "labels". The tables have been filled by phrases by the system from the conversion results, while the label table is still left empty that will require manual labeling by the researcher. The label will be entered manually based on the abbreviated name of the aspect of need. The phrase that is labeled contains the initials "other" or "O" if there are phrases not related to the label and initials from 15 aspects of the need if the phrase has something to do with the label.

3.4 Phrases List Formation

At this stage, the labeled training data will take the index of every word that contains a phrase with initials other than "other" to separate the phrases that will be stored in a new csv file named "phrase_list.csv". The phrase takes each pair of phrases and labels that have been manually inputted by the expert. The phrases in the phrase list will be sorted randomly based on the long and short-labeled words that are labeled by the researcher in the manual labeling process. In the manual labeling process, the author labels three or more words/tokens on the file, which will come out at the top of the list. In contrast, the words with fewer labels will appear in the bottom line of the phrase list.

3.5 Information Identification

In this study, the word indexing method was used to find information related to potential applicants. The information to find is the label on each file. At this stage, the system will identify the phrase by matching the label specified in the phrase_list. To carry out this process, test data is needed to determine the suitability of the created model to find the information that appears along with the names of aspects of needs. The process is identified based on the word "but" after the sentence, which will make the system classify the particular sentence as the highest need (most prioritized need) and the lowest need (least prioritized need). Afterwards, the system will check the index based on any phrases that appear in the document to identify sentences that refer to the label on phrase_list. The output will provide a description of the phrases and labels that appear on each file along with the label and a summary of the aspects of the needs that come out of the file. The txt results are stored in a different folder with the document folder. Figure 2 shows information identification process.

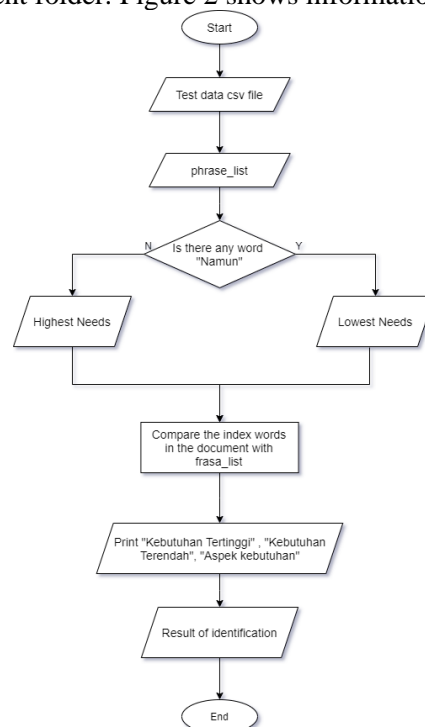


Figure 2. Information Identification process

3.6 Document Search

Once the search results of the desired aspects of needs are known at this stage, these results will become a reference to see which prospective applicants of FK UII with the highest need in that aspect. Document searching can make it easier for the organizers to find documents of prospective applicants of FK UII based on the 15 desired needs as a component of the next assessment. For example, the system can recognize the need for "ACH" as the highest need, and thus the system will display data on which applicants have the highest needs for these aspects of needs. figure 3. shows document search process

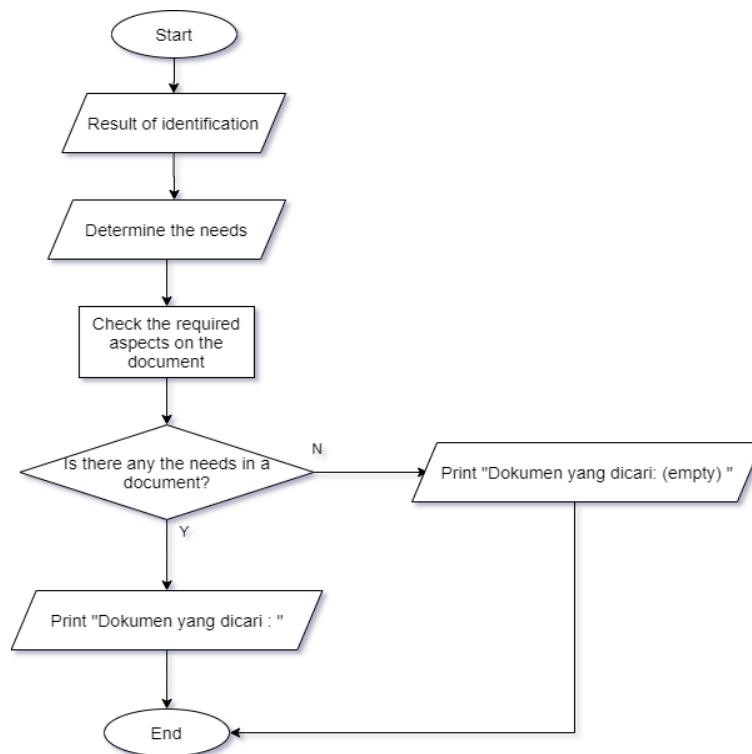


Figure 3. Document search process

3.7 Evaluation

Furthermore, to conduct a model testing, we need original data and new data to be tested will to obtain similar degree value. After obtaining a document containing relevant information, the test data will evaluate the original data against the new data by looking for the similar degree value. Then, it will evaluate each type of phrase form based on its precision, recall, and f-measure. The data to be used as test data amount to 20% of the entire data, which is equal to 0.20.

4. Result

The previous researches indicate the indexing method as a suitable method to solve the problems in document searching. The modeling stage of the obtained data aims to produce a system that is able to identify every word that appears related to 15 aspects of predetermined needs and provides information on which aspects of needs arise at the conclusion of the EPPS test. To help create this model, the researchers used the python3 programming language through the terminal and IDE Visual Code.

It is explained in formula 3.1 that the conversion process of the EPPS test results document, which is saved in the form of a pdf file into a csv file that takes the paragraph in the "description" index, will be carried out. Before converting the data into a CSV file, the text data is converted to a txt format to get all the text in the PDF file. The document conversion process requires a preprocessing process of casefolding using the `text.lower ()` function and tokenization using the `word_tokenize` function. The output of one of the 74 documents converted to documents and preprocessing is presented in the following figure.

kata	label
hasil	
pemeriksaan	
psikologis	
menunjukkan	
bahwa	
saudara/i	
maulfi	
natsir	
asyari	
memiliki	
kebutuhan	
yang	
tinggi	
untuk	
menyesuaikan	
diri	
dan	
mengikuti	
aturan/konvensi	
yang	
telah	
ditetapkan	

Figure 4. The output of document conversion

Once we obtained the converted documents, the 74 documents will be split into two parts, namely train data and test data with a default ratio of 20% of the train data. The train data and test data that have been converted will be labeled manually by the researcher by looking at several criteria from the aspect of requirements on the EPPS test. Labeling is done directly from the csv file of each document. The criteria for aspects of need along with keywords are presented in Table 1 below.

Table 1. Aspect of needs and the keywords.

Aspect of Needs	Keyword
Achievement (ACH)	Memperoleh sesuatu, mencapai prestasi, mencapai, prestasi, meraih sesuatu.
Deference (DEF)	Menyesuaikan diri, mengikuti aturan, patuh terhadap aturan.
Order (ORD)	Mengatur hidup, barang, ide, keteraturan, sistematis.
Exhibition (EXH)	Mengekspresikan diri, kepercayaan diri, yakin dengan kemampuan.
Autonomy (AUT)	Mandiri, tidak tergantung, pendirian/pendapat, tergantung pada orang.
Affiliation (AFF)	Memberikan perhatian, pergaulan harmonis, lingkungan sosial, menaruh perhatian, berinteraksi
Intracception (INT)	Berempati, bersimpati, kepekaan, memahami, merasakan, memahami orang.
Succorance (SUC)	Dimanjakan, diperhatikan, diurus
Dominance (DOM)	Mendominasi, memerintah, mempengaruhi, mengatur orang, menonjolkan diri, ketua kelompok.

Abasement (ABS)	Mengakui kesalahan, perasaan bersalah, menyalahkan diri
Nurturance (NUR)	Menolong, merawat orang/benda, menjaga barang, merawat orang.
Change (CHG)	Melakukan perubahan, menghindari rutin, sesuatu yang tetap.
Endurance (END)	Menyelesaikan pekerjaan, ketekunan, keuletan, kegigihan, ketahanan, teratur.
Heterosexual (HET)	Seksual, mengikuti pendapat
Aggression (AGG)	Mendobrak, berperilaku agresif

Furthermore, in the process described in formula 3.4, the text in the "description" is separated and stored in the csv "phrase_list" form with a table containing the pairs between the phrases and the labels. The sorting process uses the "sorted" function to sort the phrases that are labeled by authors. Phrases in phrase_list will be sorted based on phrases labeled with three or two words or only one word so that they are stored according to the longest and shortest word entered by the author. A list of the list phrases can be seen in Figure 5 below.

frase	label
yakin dengan kemampuan	EXH
sesuatu yang tetap	CHG
tergantung pada orang	AUT
patuh terhadap aturan	DEF
pengambilan keputusan	AUT
mengikuti aturan/konvensi	DEF
mengekspresikan dirinya	EXH
memberikan perhatian	AFF
merawat orang/benda	NUR
melakukan perubahan	CHG
berperilaku agresif	AGG
merawat orang/benda	NUR
mengikuti pendapat	HET
memperoleh sesuatu	ACH
pendirian/pendapat	AUT
kehidupan seksual	HET
mencapai prestasi	ACH
perasaan bersalah	ABS
menyesuaikan diri	DEF
menaruh perhatian	AFF
kepercayaan diri	EXH
tidak tergantung	AUT

Figure 5. Phrase_list

4.1. Information Identification

In the process of identifying information, the system will display the meaning contained in the EPPS test psychogram to make it easier to find out which aspects of the needs appear in the EPPS test conclusions of prospective applicants. Proses information identification we need to look for an index to separate paragraphs describing the highest and lowest needs. In this case, the word that distinguishes between the highest and lowest needs is the word "however" in the beginning of sentences.

Some characters that appear will also be replaced using the regular expression library in python. Figure 6 shows the output of the identification information based on phrase_list.csv. The phrase "Highest need" displays the phrases list containing the highest need of the EPPS test result, while the phrase "Lowest need" displays the phrase list containing the applicant's lowest need. The phrase list shows an information about the phrases and labels written in the sentence that concludes the EPPS test results. "Aspects of needs" contained in this output show briefly the aspects of needs based on the predetermined index phrases and labels.

```

1831150236.pdf
-----Kebutuhan Tertinggi-----
berempati -> INT
bersimpati -> INT
yakini dengan kemampuan -> EXH
menonjolkan diri -> DOM
ketua kelompok -> DOM
mengatur orang -> DOM
mempengaruhi orang -> DOM

Aspek Kebutuhan : INT, EXH, DOM

-----Kebutuhan Terendah-----
patuh terhadap aturan -> DEF
patuh -> DEF
aturan -> DEF
mengikuti pendapat -> HET
berperilaku agresif -> AGG
agresif -> AGG
mencapai prestasi -> ACH
mencapai -> ACH
prestasi -> ACH
memperoleh sesuatu -> ACH
sesuatu yang tetap -> CHG
rutin -> CHG
ketahanan -> END
kegigihan -> END
ketekunan -> END

Aspek Kebutuhan : DEF, HET, AGG, ACH, CHG, END

1831150091.pdf
-----Kebutuhan Tertinggi-----
mandiri -> AUT
tidak tergantung -> AUT
tergantung pada orang -> AUT
pendirian/pendapat -> AUT
menaruh perhatian -> AFF
bergaul -> AFF
harmonis -> AFF
menolong -> NUR
merawat orang -> NUR
benda -> NUR

Aspek Kebutuhan : AUT, AFF, NUR

-----Kebutuhan Terendah-----
ketahanan -> END
kegigihan -> END
ketekunan -> END
mencapai prestasi -> ACH
mencapai -> ACH
prestasi -> ACH
memperoleh sesuatu -> ACH
patuh terhadap aturan -> DEF
patuh -> DEF
aturan -> DEF
mengikuti pendapat -> HET
keteraturan -> ORD
sistematis -> ORD
kehidupan seksual -> HET

Aspek Kebutuhan : END, ACH, DEF, HET, ORD

```

Figure 6. An example of a document resulting from the information identification process

4.2. Document Search

Once the system finds and identifies the information, it will be easier for us to search for documents through the aspects of the needs listed in the results of each document that was identified in the previous process. This document search can be done everytime a user wants to find an information from multiple documents based on which of the 15 aspects of the desired need is required. If the user requires the search for "ACH" and "DEF", the system will randomly search for the document containing those words as the highest aspect of needs according to the desired result. Figure 7 shows the results for documents search process using the desired label of "ACH" and "DEF" which can be stored in the form of a txt file as what the experts required.

```

-----Dokumen yang dicari : -----
1831150091.pdf
1831150239.pdf
1832150031.pdf
1832150046.pdf
1831150104.pdf
1831150115.pdf
1832150084.pdf
1832150073.pdf
1832150027.pdf
1831150207.pdf
1831150214.pdf
1831150002.pdf
1831150202.pdf
1831150230.pdf
1831150177.pdf
1832150022.pdf
1832150069.pdf
1831150020.pdf
1832150091.pdf
1831150108.pdf
1831150090- Inkonsisten 9.pdf

```

Figure 7 The results of file implementation based on aspect of needs "ACH" and "DEF"

4.3. Evaluation

The next stage is to test and evaluate the test data with the original data on the system to look for the appearance of similar degree values. It showed 14 test data that had been split previously in the 3.2 process. In this evaluation, precision, recall, and f-measure are used as a form of evaluation of the system. The final results of the model testing are resulted from the calculation of the average value of each document shown in **Table 2**.

Table 2 The average per documents

	Precision	Recall	F-Measure
Mean	0,89	0,93	0,91

Table 2 displays the average results of the evaluation using different types of phrase forms. This search obtained the best results of all reduced phrase form types with an f-measure value of 0.97 for all data. The above test result does not contain the aspect of accuracy because accuracy is more appropriate for class distribution, while f-measure is a better metric for cases with an imbalance in evaluating the model.

5. Discussion

In this study, the system was designed using an indexing method based on a list of phrases the researcher created to obtain information in the text "description" of the EPPS test results. Based on the literature review, the indexing method in this study is done by adding a list of phrases as a dictionary to help identify information and to find out which documents are needed by the organizer. Figure 3 displays 60 lists of phrases that are formed. Phrase list formation is done by looking at examples of cases of phrase truncation and name entity recognition, where the type of truncation is based on the phrase tag used in the parse tree and the name of the phrase tag and the method of classifying text phrases with certain rules based on word classes. (POS) [10], [11].

The indexing method obtained a quite good result since it only used small amount of data. These results are obtained from a list of identified phrases, which are used as a reference. As a result, the system can match the information that appears in the EPPS test "description" text by producing phrases and labels along with the appearing aspects of needs to ease users in reading the information. However, such method is still constrained when it has to deal with a huge amount of data because it will need a thorough update to check the accuracy of manual labeling. Nevertheless, since this case used manual labelling, it is necessary to add an updated system to check system performance to avoid any errors in the inputted data. Therefore, it is pivotal that an expert in this field does manual labelling to avoid misinformation.

In this case, the system does not focus on using stopword removal, but on the removal of the word "other" contained in documents that have been manually labeled "O" to create a list of phrases or phrase_list. The development of stopword removal is a considerable topic for future research to make sure whether it should be used or not. In the evaluation, the results are not counted into the calculation of accuracy because in the search results the accuracy will be close to 100% or the "True Negative" result, which has an infinite value although in reality the results are just the reverse. Some existing examples, as shown by the TN, contain words that cannot be identified and should not be identified, but are still calculated using an accuracy formula that produces results close to 100%.

6. Conclusion

Based on the initial stages until the final stage of the research, we can conclude that the method of identifying information through the indexing process is an applicable method. This method can be

used to find information related to the aspects of the needs required by the structured EPPS test for prospective applicants, since it can search for information by referring to the information identified in the `phrase_list`. The f-measure calculation obtained the result of 0.91, the precision of 0.89, and the recall of 0.93.

However, this method is not a benchmark to create a good system, since it still needs further updates and development. Therefore, it is expected that there will be further research to develop the current research with some improved elements such as by increasing the number of datasets and continuously updated list of phrases in order to get better results. It is also necessary to develop the current research by constantly checking and correcting keywords in the converted PDF file. Other possible developments can take the form of using other more suitable methods out of the indexing method.

References

- [1] Hidayat W 2013 *Indexing and Retrieval Engine untuk Dokumen Berbahasa Indonesia dengan Menggunakan Inverted Index*
- [2] Heriyanto 2015 *Pencarian Dokumen Teks Arsip Surat* **2015** pp 173–177
- [3] Broder A. Z, Eiron N, Fontoura M, Herscovici M, Lempel R, McPherson J, Qi R and Shekita E 2006 *Indexing Shared content In Information Retrieval Systems* **3896** pp. 313–330
- [4] Haggard L, Peterson S, Rhodewalt F, Charles P, Larsen R.M and Matarazzo J.D 1983 *References Psychological Test Usage Patterns in Five Professional Setting Kansas City* pp. 857–861
- [5] Piedmont R.L and Costa P 1992 *An Assessment of the Edwards Personal Preference Schedule From the Perspective of the Five-Factor Model*
- [6] Trisna I. N. P and Nurwidyantoro A, *Single Document Keywords Extraction In Bahasa Indonesia Using Phrase Chunking* **18** p. 1917
- [7] Boubekeur F and Azzoug W 2013 *Concept-Based Indexing in Text Information Retrieval* **5** pp. 119–136
- [8] Lu, Guajun, *Multimedia Database Manajemen Systems*, 1999 Artech House, Inc
- [9] Bakala, *Information Retrieval System By Using Vector Space Model*, 2019,p.1563
- [10] Yang J 2004 *Phrase Chunking for Efficient Parsing in Machine Translation* **2972**, pp. 478–487
- [11] Arman A. A , Putra B. Arif N, Purwarianti A and Kuspriyanto 2013 *Syntactic Phrase Chunking for Indonesian Language* **11** pp. 635–640
(Bakala, 2019)