Web-Based Application of Indonesia-Manado Translation Forum Using Extreme Programming Methodology

Vivi Peggie Rantung¹, Abimanyu Marvie Dwisuprapto*², Ferdinan Ivan Sangkop³
¹,²,³Teknik Informatika, Fakultas Teknik, Universitas Negeri Manado, Tondano
e-mail: ¹vivirantung@unima.ac.id, *²18210008@unima.ac.id, ³ivan.sangkop@unima.ac.id

Abstract
Manado language (Manadonese) is one of the languages that is registered in the ISO 639-3 standard system with the code of xmm and used by 3.320.000 users. In daily life, Manadonese is commonly used verbally. In its oral use, Manadonese becomes a low-resource language, meaning it lacks text-based resources, which makes it difficult to develop various linguistic-based technologies to preserve Manadonese. This research aimed to provide a data source to develop advanced linguistic-based technology. The methodology used for this research is extreme programming (XP). The result of this research is a Web-Based Application Of Indonesia-Manado Translation Forum with various functions, such as uploading Indonesian articles that will be translated into Manadonese, from this diverse amount of translation the best of it will be chosen to be processed every word and sentence that will be saved in the database. The application is tested with Acceptance Testing as an indicator and the result shows that the forum managed to accomplish the goal set up before.

Keywords— Web Application, Forum, Indonesian, Manadonese, Extreme Programming

1. INTRODUCTION

Humans in their long life are social creatures that need intense interaction with other humans, this intense interaction can also be called communication, to understand each other when communicating humans invented language that is used to attain some communicative intent, and there is more than one type of communicative intent, sometimes it is to convey information to other people, or simply for socializing, or even to ask someone to do something to us[1]. Language as a communication “tool” is used by every human in this world, including the people of North Sulawesi Province in the Republic of Indonesia, especially in the region of Manado City and the area near it. The language that is used in this particular region is Manadonese, which is registered with the code xmm in the ISO 639-3 standard system, and according to the census, Manadonese used by 3.320.000 speakers, which 1.820.000 of it are L1 Users and 1.500.000 are L2 Users[2].

Even with the statistics mentioned above, Manadonese is a language with very little amount of data, which makes it very hard to study or to be computerized and can be hard to directly process with the statistical method for the very reason, this problem makes Manadonese can be defined as low-resource language[3]. The reason for this lack of data can be traced to the lack of written text in Manadonese because this language is mostly used verbally, and because of this, Manadonese can be under threat of extinction. Seeing this can be understood that to preserve Manadonese as a language the action of curating text-based data is a must, where this text data can be used in the Map and Registration of Online Language and Literature program as a means to preserve and protect this language[4].

To ensure the development of the application runs smoothly then it needs to be done using a system development life cycle (SDLC), which in this research uses the extreme programming methodology, this method has an advantage that can fasten the development of a project because in its implementation this method usually only required a small number of
people in a team that can rapidly process every change in requirements that occurs in production[5], and this research[6] conclude that extreme programming can produce software in such a short time and with a fulfilled requirements, implementing extreme programming keep the development in such a simple way and keep all the related party to actively involved in the development so that other than the software be done quickly but also keep on fulfilling its original or revised requirements[7].

While extreme programming is used as SDLC the technology used to get the data from the developed forum, the web scraping technique is used, this technique uses HTML protocol to fetch data, extract it and finally transform the remaining data to the form of structured format that can be saved to be used later[8] and to process the translation that been submitted by the public user the technology used are nltk library of python language that can process text very quick and easy[9] using nltk can ease the process of tokenizing every word of the translation submitted.

Based on the states of Manadonese as a language that has been explained above, this research has a goal to develop a platform with a means to collect data, with the said platform is a Web-Based Application Of Indonesia-Manado Translation Forum that is developed by implementing extreme programming as its SDLC. By using this application, the text-based data necessary to be collected to enrich Manadonese as a language, especially in terms of its literature sources. The developed application can scrap news articles, upload them, and accept any translation of the said article, other than that the system can process the translation word-by-word to be compared with the existing Indonesian language and can determine which word it should save as Manadonese data text. Hopefully, the data collected can be used in further research for more advanced language-based technology, especially using Manadonese as the base language.

2. RESEARCH METHOD

The web-based application of this research is developed by implementing extreme programming as its SDLC this method is a part of the Agile Software Development approach that has a good advantage in its effectiveness, efficiency, and flexibility[10], also this method is used because of its fundamental value that in the process of development showed the ability to respond to any change in requirements or any other possibility that might happen, this value is communication, simplicity, feedback and courage[11].

By doing a literature study of various types of research that implement the extreme programming methodology[7],[12] the development process of this research will be done based on this research, and following this, the steps done to develop this web-based application are planning, designing coding, and testing.

![Figure 1. Extreme Programming Method](image)

2.1. Planning

Planning is the first step in this whole research, on this stage some initial research is conducted, to identify the problems, and analyze the web-based application requirements.
Observation of the environment is the first step in doing the initial research, considering the nearby society uses Manadonese as their lingua franca this step is a must to identify their way of communication and consider what can be implemented in the web-based application based on the observation results, other than doing observation an interview also conducted to people that used Manadonese.

Observation and interview are a means of collecting data and are done to understand how they can translate an Indonesian article to Manadonese and what kind of feature shall be implemented in the end product. The results of the collecting data steps are processed into the form of user stories so the requirements can be easier to understand, this user story has its template form to ease the process of creating it, the template that will be used is “As a [Who] I want to [What], so that [Why]”[14] this user stories later will be converted into features needed for the web-based application.

2.2. Design

Requirements acquired from the planning stage are processed into system design, this stage is done to design the model of the web-based application project in the form of a logical model that explains how the application will operate after the development is finished[15], other than that converting user stories to a model can make helped others to understand how the system work easily.

In this research the modeling is done using a use case diagram, where the diagram is used to explain the active aspect that happens on the web-based application[16], and the entity relationship diagram is used to design the database, where the data will be represented visually to show how the relations between each entity on the database[17], both of this model will help when the time came to implement the code for every feature planned.

2.3. Coding

Based on the requirements and the model that had been defined before, the next stage is the coding where all the results of the previous steps will be implemented into the web-based application by doing a coding session, in this research the language used for the coding is the Python programming language and using Django as its web framework.

Every coding session will have finished features that will be released and tested by the proposed user, and by following one of the fundamentals which is feedback, the code will be reviewed and altered if necessary after receiving feedback from the user based on their experience using the released features.

2.4. Testing

After every release, a test will be conducted, to make sure that the released features fulfill the acceptance criteria, and as a prerequisite for advancing to the next coding session, if the released features satisfy the criteria then the coding session is a success and can advance to do another coding session for the next features.

The tests that will be done in the development process of this research are the acceptance testing and the usability test, where the acceptance testing is conducted based on the criteria that had been made before to see if the web-based application already satisfied the requirement according to the proposed design[18] and the usability test are used to see if the user can effectively use the web-based application as it should be[19].

3. RESULT AND DISCUSSION

This research will have an end product of a web-based application that can provide a forum so that people can translate Indonesian news articles into Manadonese and collect the words and sentences that are in Manadonese by processing this translation. The web-based application is developed using Django web framework that is built using Python programming
language and MySQL as its RDBMS, the reason Python is used is because of its vast amount of libraries that can process text, namely nltk, and beautiful-soup are the primary library used for collecting data process in this web-based application.

The steps taken to develop this web-based application are as follows.

3.1. Planning

The first stage of planning to develop this web-based application is collecting data by doing observation and interviews, the observation process is to understand how the nearby societies communicate with each other and to translate it into a feature, other than observation, interviews have also been conducted to find out what kind of feature that should be included in the project.

The result of observation and interview was later converted into the user stories table as shown in Table 1, this user stories use the template form of “As a [Who] I want to [What], so that [Why]”.

<table>
<thead>
<tr>
<th>Table 1. User Stories</th>
</tr>
</thead>
<tbody>
<tr>
<td>As a</td>
</tr>
<tr>
<td>Administrator, Verifier, User</td>
</tr>
<tr>
<td>Administrator, Verifier, User</td>
</tr>
<tr>
<td>Administrator</td>
</tr>
<tr>
<td>Administrator</td>
</tr>
<tr>
<td>Administrator</td>
</tr>
<tr>
<td>Administrator</td>
</tr>
<tr>
<td>Administrator</td>
</tr>
<tr>
<td>Administrator</td>
</tr>
<tr>
<td>Administrator</td>
</tr>
<tr>
<td>Administrator, Verifier, User</td>
</tr>
<tr>
<td>Verifier</td>
</tr>
<tr>
<td>User</td>
</tr>
<tr>
<td>User</td>
</tr>
<tr>
<td>User</td>
</tr>
<tr>
<td>User</td>
</tr>
<tr>
<td>User</td>
</tr>
</tbody>
</table>

Based on Table 1 can be identified all the roles that will interact within the system, these roles are ADMIN, VERIFIER, and USER. The role of an ADMIN will be assigned to the webmaster, VERIFIER will be assigned to the party that can evaluate how good is the translation, and USER is the general user that can access the forum where they can either translate an article or just merely read other user translation.

3.2. Design

After collecting data and converting it into the user stories table, the development is advanced to the next stage which is to design the features of the system as a model in the form of a use case diagram and entity relationship diagram, with the use case as shown in Figure 1 is how the dynamic between actors inside the systems and figures 2 shows the relations between all the table in the database.
Figure 2. Use Case Diagram

Figure 2 above shows a use case where the three roles described in the planning process are defined as actors, which means all the actors involved are Admin, Verifier, and User. The Admin can create a post of a news article, or announcement, close a post, and manage the Verifier, the Verifier will have access to evaluate user translation and the User can enroll as Verifier, also view and translate a post.

Figure 3. Entity Relationship Diagram

To accommodate the features that involve collecting and processing data, the web-based app will need a database to store the data, the RDBMS used is MySQL. In Figure 3 the relationship between the table in the database is visually represented in the form of an entity relationship diagram, the relations involved are one-to-one, one mandatory to many optional, one mandatory to one optional, and one mandatory to many mandatory.

3.3 Coding

After finishing the designing process the next step is coding, this coding session will be heavily influenced by the model before and the feedback that came after the release. This coding
session started with writing the code in the Django web framework and migrating the database created to the MySQL RDBMS.

Then after finishing a coding session a small release will be done and feedback is expected, if the small release is satisfactory to the requirements set before then the next coding session will begin, and if the feedback provides new requirements or there is an error that must be fixed then the code session will begin again to review and rewrite the code if necessary.

```python
def compare(translasi):
    with open("main/idwords.html", "r") as file:
        body = file.read()

        soup = BeautifulSoup(body, "lxml")
        word = soup.select_one(selector=".word").get_text(strip=True)
        allwords = word.split()

        text = translasi
        regex = re.compile("[^a-zA-Z\s]", " \t",
                          re.IGNORECASE)

        low_text = [(lowtext.lower() for lowtext in text)
                    if lowtext
                    else []
                   for lowtext in text]
        low_allwords = [(lowallwords.lower() for lowallwords in allwords)
                        if lowallwords
                        else []
                       for lowallwords in allwords]

        clean = sorted([set(low_text)])
        allword = sorted([set(low_allwords)])
        compared = sorted(list(set(clean)) - set(allword))

        context = {'mdo': compared}
        return context
```

```python
def compare_db(kata, kata_db):
    clean = sorted([set(kata)])
    allword = sorted([set(kata_db)])
    compared = sorted(list(set(clean)) - set(allword))

    context = {'mdo': compared}
    return context
```

```python
def kalimat_perbandingan(ind, mdo):
    tokenizer = RegexpTokenizer(r"[,\(\)]+")
    text_compiler = re.compile("\'<\*.*?>\'")
    ind = re.sub(text_compiler, "", ind)
    mdo = re.sub(text_compiler, "", mdo)

    kalimat_ind = list(map(str.strip, tokenizer.tokenize(ind)))
    kalimat_mdo = list(map(str.strip, tokenizer.tokenize(mdo)))

    context = {'ind': kalimat_ind, 'mdo': kalimat_mdo}
    return context
```

```python
def kalimat_mdo(mdo):
    tokenizer = RegexpTokenizer(r"[,\(\)]+")
    text_compiler = re.compile("\'<\*.*?>\'")
    mdo = re.sub(text_compiler, "", mdo)

    kalimat_mdo = list(map(str.strip, tokenizer.tokenize(mdo)))

    context = {'mdo': kalimat_mdo}
    return context
```

**Figure 4. Post page with close post button**

The code in Figure 4 above is the function that will support the code below in Figure 5, this function will use beautiful-soup and nltk library to do a data scraping and text tokenizing to process and divide the Manadonese word, allowing it to be stored in the database and below in figure 5 shows what the system run when processing the translation data.

```python
translasi = translasi.content
hasil_banding = compare(translasi)
hasil = hasil_banding["compare"]

kata_db = Manado.objects.all().distinct()
list_kata = []
for kata_db in kata_db:
    kata_db = str(kata_db)
    list_kata.append(kata_db)

banding_kata = compare_db(hasil, list_kata)

if banding_kata == None:
    pass
else:
    for hsl in banding_kata:
        Manado.objects.create(kata=hsl)

kalimat = kalimat_perbandingan(post_konten, translasi)
kalimat_ind = kalimat["ind"]
kalimat_mdo = kalimat["mdo"]

zipped = zip(kalimat_ind, kalimat_mdo)
for ind, mdo in zipped:
    Perbandingan.objects.get_or_create(kalimat_ind=ind, kalimat_mdo=mdo)
```

**Figure 5. Process translation code**
3.4 System Implementation

The result of the coding session is a release with a finished feature. This section will show the release that includes a major feature in it the first as shown in Figure 5 below.

![Figure 6. Forum homepage](image)

Figure 6 is the homepage that every actor will meet the first time they access the forum, this page has a few statistics.

![Figure 7. Upload news article post](image)

Figure 7 above shows how the Admin uploads a post by inserting a news article URL, the beautiful-soup library can fetch the data and after that, the Admin can confirm to upload the post. After the post is uploaded, the Admin can also close the post so that it can be evaluated by the Verifier, how the Admin can do this is shown in figure 8.

![Figure 8. Post page with close post button](image)

Figure 8 shows a button that the Admin can use to close the post status and can finally be evaluated by the Verifier.

![Figure 9. Post announcement form](image)
Figure 9 above is the page that will be used by the Admin to write and post an announcement, this type of post can be used by the Admin to announce a change in the forum or another type of announcement.

Other than uploading a post, the Admin also can manage the existing Verifier and add a new one. Figure 10 shows the page that contains all the Verifiers that existed, and also has a button to add a new Verifier, this button will show a form to add a Verifier as shown in Figure 11.

The form shown in Figure 11 is to be filled by the Admin according to the new Verifier data, this Verifier can be an individual or an organization.

Previous features are what the Admin can do, figure 12 shows a page that has the feature the Verifier can do, this page contains an evaluating form that will be filled by a Verifier.
Not only evaluating the translation, the Verifier can also leave an evaluation message on a form as shown in Figure 13 above, this will allow the User who translates the article to understand why they got the points on their translation.

![Figure 14. Translate Article Form](image)

The user, as shown in Figure 14 can view and translate the news article, this translation is the one that will be evaluated by the Verifier and the data will be collected to be processed and stored in the database.

![Figure 15. Request Article Form](image)

Other than viewing the post and translating it user can also request an article, to do this the User can access the form on the page shown in Figure 14 above, by requesting the article user can ask for their preferred type of article.

![Figure 16. List of Manadonese words collected](image)

Figure 16 above shows the data of Manadonese words managed to be collected in the form of word-by-word.

### 3.5. Testing

After the implementation is successful the next stage is testing the forum. The test conducted is an acceptance test and usability test, where the acceptance test is to see if the finished forum application already fulfilled the satisfactory standard by the criteria set before and the usability is to see if the user can use the forum application as its intended.

The acceptance test results based on the criteria sets are shown in Table 2, this table will contain features, acceptance criteria, and results.
The technique used to calculate the results is the performance measurement technique[20] and the formula [7] used is as shown in Formula 1.

<table>
<thead>
<tr>
<th>Features</th>
<th>Acceptance Criteria</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Login &amp; Logout</td>
<td>Access the forum according to the account role and its authorization and log out from the forum.</td>
<td>Success</td>
</tr>
<tr>
<td>Register</td>
<td>Create a new account in the forum</td>
<td>Success</td>
</tr>
<tr>
<td>Register as a Verifier</td>
<td>Registered users submit a Verifier application</td>
<td>Success</td>
</tr>
<tr>
<td>Manage Profile</td>
<td>All actor can manage their profile, such as changing their profile picture or name.</td>
<td>Success</td>
</tr>
<tr>
<td>Upload URL</td>
<td>Receive news article URLs from the domain of CNNIndonesia.com, Tempo.com, Kompas.com.</td>
<td>Success</td>
</tr>
<tr>
<td>Scrap URL content</td>
<td>Get the news article data contained in the uploaded</td>
<td>Success</td>
</tr>
<tr>
<td>Upload URL content as an article post</td>
<td>Store the content from the URL to the database</td>
<td>Success</td>
</tr>
<tr>
<td>Close news article post</td>
<td>Update post status to False or Closed</td>
<td>Success</td>
</tr>
<tr>
<td>Upload announcement post</td>
<td>Upload and store new posts with the type of announcement</td>
<td>Success</td>
</tr>
<tr>
<td>Update announcement post</td>
<td>Update the announcement content with the new data input</td>
<td>Success</td>
</tr>
<tr>
<td>Scrap and process the data from the submitted translation</td>
<td>Get the translation with the highest point and run a split(), tokenize(), and compare() function then save the processed data</td>
<td>Success</td>
</tr>
<tr>
<td>View User</td>
<td>Get data of all Users and Verifier and display it to the Admin User list</td>
<td>Success</td>
</tr>
<tr>
<td>Add new Verifier</td>
<td>Register a new verifier account through the form</td>
<td>Success</td>
</tr>
<tr>
<td>View all Manadonese words.</td>
<td>Get and display all the Manadonese words stored in the database</td>
<td>Success</td>
</tr>
<tr>
<td>Download Manadonese sentences and words.</td>
<td>Get all the Manadonese words and sentences and download them onto the local device in an excel document.</td>
<td>Success</td>
</tr>
<tr>
<td>Translate post</td>
<td>Get and Store User translation input according to its related post</td>
<td>Success</td>
</tr>
<tr>
<td>Evaluate translation</td>
<td>Get and store Verifier evaluation input according to its related</td>
<td>Success</td>
</tr>
<tr>
<td>Accept and accumulate points of submitted translation</td>
<td>Get the point from translation and accumulate the current points of the User</td>
<td>Success</td>
</tr>
<tr>
<td>Request article</td>
<td>Request news article request by sending the URL</td>
<td>Success</td>
</tr>
<tr>
<td>Process the requested article.</td>
<td>Process and upload the content of the valid requested URL</td>
<td>Success</td>
</tr>
</tbody>
</table>

The results from Table 2 above show that the finished forum can fulfill the criteria and requirements set before and has satisfactory functionality.

The next test is the usability test, which will measure the effectiveness of the forum. The results will show how easily the forum can be used by the user. This test was conducted by 4 users following the scenario shown in Table 3.

<table>
<thead>
<tr>
<th>Features</th>
<th>Codes</th>
<th>Scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manage verifier</td>
<td>S1</td>
<td>You are an administrator, provided with a list of eligible verifiers, and use the system to add new verifiers or remove them.</td>
</tr>
<tr>
<td>Upload news article post.</td>
<td>S2</td>
<td>You are an Admin who wants to make a new post, provided with the URL from the eligible domain. Use the system to post a new post.</td>
</tr>
<tr>
<td>Upload announcement</td>
<td>S3</td>
<td>You are an Admin and need to announce all the forum members. Use the system to post a new announcement.</td>
</tr>
<tr>
<td>Translate news article posts.</td>
<td>S4</td>
<td>You are a User who wants to translate some articles into Manadonese. Use the system to translate the article.</td>
</tr>
<tr>
<td>Evaluate news article translation.</td>
<td>S5</td>
<td>You are the Verifier that needs to evaluate some translation. Use the system to evaluate the translation.</td>
</tr>
</tbody>
</table>

The scenario in Table 3 above is defined so that it can be more effective when the test is conducted. The users involved in this test are public users that able to use the internet especially those who can use Manadonese as their first language, there are 5 users chosen to participate in this test who can use the internet and are proficient in Manadonese as it is their main language and also the chosen participant are fellow student who are not particularly experienced in this type of forum either as administrator, verifier or as a translator. This type of participant is chosen to make sure that the forum is easy to use by all upcoming users as long as they can use the internet. The participants are picked to test the usability of the forum interface according to the scenario in Table 3. The technique used to calculate the results is the performance measurement technique[20] and the formula [7] used is as shown in Formula 1.
Effectiveness = \frac{i}{n} \times 100\% 

(1)

Where \( i \) is the task that was done successfully and \( n \) is the total of all the tasks, in this case, if the user manages to do a scenario successfully then the results will be given a point of 1, and if failed then the point is 0. The test is conducted and the results of it are as shown in table 4.

Table 4. Final Effectiveness Test Result

<table>
<thead>
<tr>
<th>Features</th>
<th>Scenario</th>
<th>U1</th>
<th>U2</th>
<th>U3</th>
<th>U4</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>S2</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>S3</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>S4</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>S5</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Effectiveness Average</td>
<td></td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 4 shows the performance measurement technique produces a 100% rate of effectiveness, by the results its produce, it show that the forum is easy to use.

4. CONCLUSION

Developing a web-based application by implementing extreme programming can ease the development phase and produce a quality end product which in this case is the Web-Based Application Of Indonesia-Manado Translation Forum that by the acceptance criteria is fulfilling all of its purposes, especially collecting text data of the Manadonese, the usability testing proof that the product is can be use by the public user easily in particular of those who can use the internet and proficient in Manadonese.

5. SUGGESTION

Based on the research, the author suggests another feature to further clean the collected data so it can be more effective to use in future research also a simple dictionary of Manadonese can be implemented using the existing Manadonese word so user’s can understand more of certain Manadonese word and its meaning in Indonesian.

ACKNOWLEDGEMENTS

This research was conducted with the help and support of Manado State University where the authors are part of the institution as a student and lecturers.

REFERENCE


