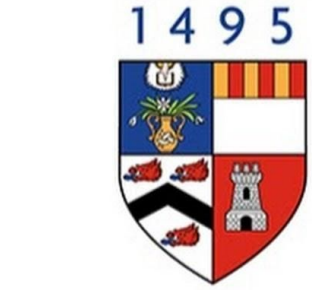


UNIVERSITY OF
ABERDEEN

CS551H:
Natural Language Generation
Assignment III

51986424 - Abdelrahman Morad
51987820 - Avinash Bagul
51985867 - James O'Donoghue
51986376 - Suryansh Bothra
Authors: **51988659 - Zhiwei Lu**

Revision Date:
15th March 2020

Table of Contents

Part 1 – Commercial Analysis

We have divided the commercial analysis of the product in the following sections which will cover the key areas of SDLC and help us to develop and track the software cost.

Part 1.1 – Vision

The vision of this application to achieve below the following Use Cases:

1. To create an automated end to end forex reporting for 3 currencies or more considering the baseline currency to be USD.
2. Reports will be generated by collecting Forex data from renowned sources.
3. After cleansing and preprocessing the data would be catered to Arria NLG to generate the report.
4. The generated text will be fed to Style Transfer module which will transform the report into a customized style type.
5. We will try to integrate this system to virtual assistants like Siri and Amazon Alexa in future.
6. We are using machine learning in WEKA to generate the predicted values for the currency, in future we will be adding all above transformations to convert it into speech.
7. Furthermore, we have added speech functionality to convert the generated report into speech.
8. As the system is multilingual it can be implemented in various regions and can be podcasted in FM frequency.

Part 1.2 – Market

The Forex market has been especially alive and active since the 1970's (Violeta Gaucan et al, 2010). It is the world's largest financial market. The Forex industry is global, with online traders being based globally. Nowadays all that's required is an internet connection for forex users to begin participating in the markets. There are 4.5 billion internet users in the world (Internet World Stats, 2020) and roughly 10 million forex users, which makes 1 in roughly every 450 an online trader (Broker Notes, 2017).

Fig 1: Global Forex Traders

This shows just how much of a demand there is for an end-to-end Forex reporting NLG system like the prototype that has been created by the group. Even if only 3% of the Forex trades use our system, that's still potentially ~300,000 people that might invest in our NLG system.

Bloomberg offers Forex analysis, through their *Bloomberg Terminal*, at roughly ~$20,000/year for all financial data analysis, even if you only want Forex analysis, you still have to pay for their extra financial analysis (Zachary M. Seward, 2013). With our system, we are only focusing and marketing at forex traders. We believe an annual fee of ~$3,000/year would be applicable for our end-to-end NLG system.

Our system market benefits:

1. Easy Access – Available online
2. Access via smartphone/tablet/PC

3. Usability to users with non-financial/forex terminology. Don’t need to know finance as the Arria narrative uses layman's terms.

Part 1.3 – Competitors

According to (Neil, 2003), Forex is by far the largest market in the world and it continued to hold that place ever since, with market cap averaging $5.1 trillion per day (Venketas, 2019) it's not surprising to find a lot of big and small companies directed at many aspects of this market such as trading, providing analysis, trading plans, trading platforms, among many others.

We focus on competitors who provide technical analysis as the main competitors for our product since the product doesn't consider fundamental analysis.

There are two types of websites

1- Sites who offer free technical analysis, reports and interactive charts. These sites make money in two basic ways: selling subscriptions and selling advertising (Gittler, 2018). These sites, however, don't offer customized analysis. If a trader is interested only in high-frequency trading or weekly trading he might or not find a report that suits their needs and must rely on his knowledge to do the analysis himself. This type relies on selling advertising for a profit.

2- Sites that offer customized technical analysis based on customer needs (MatrixTrade, n.d.), through a monthly subscription. Their reports are detailed with charts, technical evidence and text suggesting a trading strategy. These sites typically charge ~$100/month

Part 1.4 – Cost

Development of this software depends on its complexity, product size, development time frame and the technology used. The proposed architecture combines several technologies such as NLG, Machine Learning algorithms, Weka for prediction, voice-oriented output and UI/UX (based on ML with personalization) together.

Factors affecting Cost:

- Use of high level and latest trending technologies (NLG, Machine Learning Algorithms).
- Team of experienced engineers, £600-900/day per engineer.
- The Agile Development model was used for development.
- Time-Frame for the full project development was set to 8-9 weeks excluding weekends i.e. 40-45 working days.
- Testing – it was planned to be done with a team of experienced testers and also by launching a beta version of the application.
- Maintenance cost – Any change in the main report format or addition of new calculations and features will have standard development charges as per the time required, addition of new technology and engineer cost.
- Platform – We plan to make the software compatible for Windows, Mac, Linux, Android, IOS and Web.

External Factors affecting Cost:

The prototype uses following services with limited and free trial access to the service.

- Data – Data is collected from a paid website fixer.io (Fixer, 2020).
- Voice oriented service uses voice from Microsoft azure which again is a paid service.

Discussed above are the factors affecting the cost of our Forex Expert software. Some other costs include marketing the product through a proper channel targeting the desired customers and clients. The overall cost for the system will be a bit expensive due to use of high end technology and making it available for popular platforms along with deployment of API. However, we have considered some additional overheads at each stage.

Ref #		Week 1-3	Week 4-6	Week 7-9	Amount
1.1	Planning Phase	£3,000	£3,000	£1,000	£7,000
1.2	Requirements Phase	£2,000	£2,000	£1,000	£5,000
1.3	Development Phase	£5,000	£10,000	£20,000	£35,000
1.4	Testing Phase	N/A	N/A	£5,000	£5,000

1.5	Implementation Phase	N/A	N/A	£3,000	£3,000
				Total	**£55,000**

Part 1.5 – Why people will buy our product

A good natural language generation software needs to be able to meet the following challenges: evaluation challenge, vague language challenge, narrative challenge and communicating data quality (Reiter, 2019). The first challenge is whether NLG can convince users. This can be compared with non-NLG method. The second challenge is to compare the use of fuzzy language in two ways. People prefer qualitative terms. Using fuzzy language can make people understand the content better (Van Deemter, 2012). The third challenge is to compare the text structure generated by the NLG method. This challenge does not need to be compared with the non-NLG method, and will not be described here. The last challenge is based on the problem of incomplete information. This challenge is the same for both implementations and will not be described. Here we only compare the text generation part of our product with the charts generated by non-NLG technology. The charts generated by non-NLG technology are generated by python code based on the data of the last year. The NLG technology used is implemented by python combined with arria studio. The result of the line chart generated by non-NLG technology is shown below.

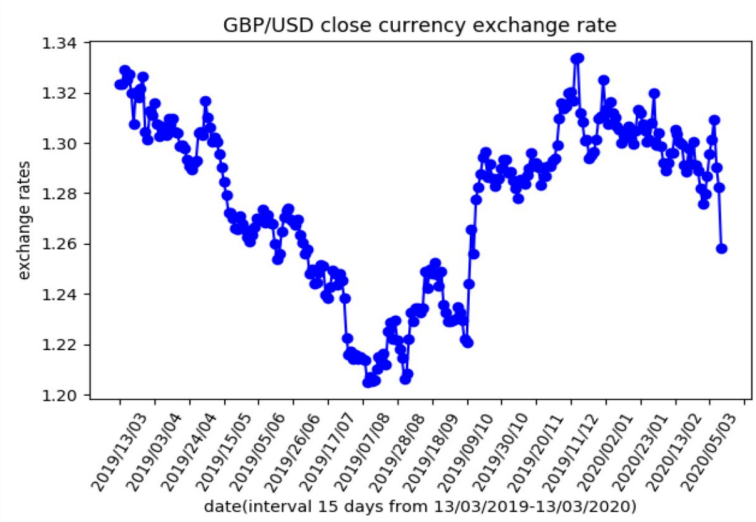


Figure 2: The latest year's exchange rate trend of the British pound to US dollar

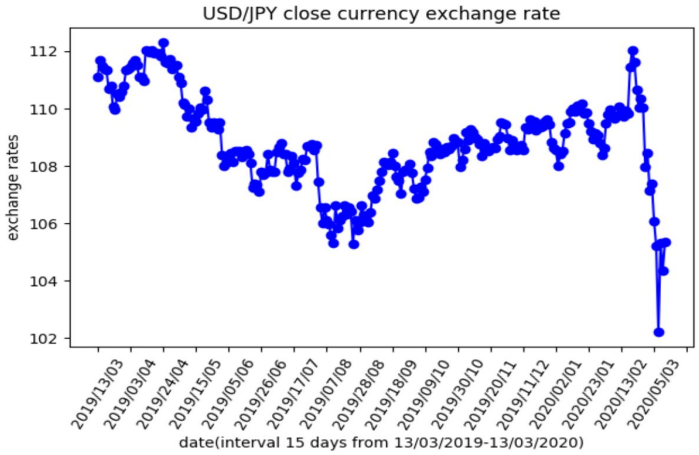
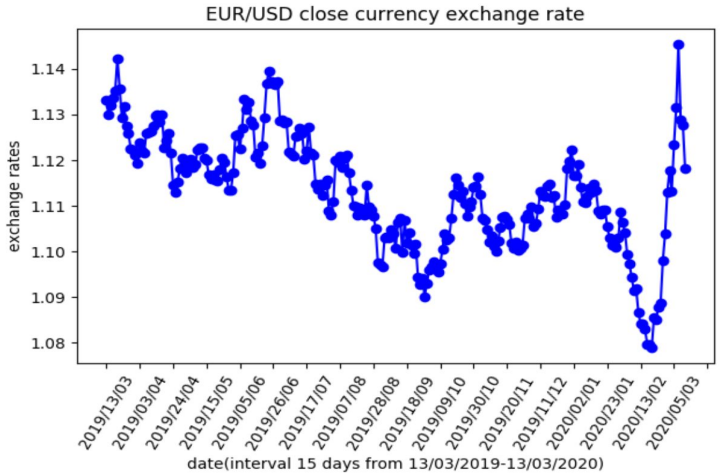


Figure 3: The latest year's exchange rate trend of the US dollar to the Japanese yen



Euro to USD exchange rate trend in the latest year

The text content generated by NLG technology is shown below.

Preview the output for row 1

Input Row:

« First ‹ Prev 1 Next › Last »

currency	isLastWeekHigher	isContinues	isHigherThan7WeekAvg	moreVariance		
EURUSD	FALSE	TRUE	TRUE	TRUE		

Output Text:

The exchange rate of Euro to US Dollar has declined in the last week compared to the previous week. This is another fallen after the previous week. Compared with the average of the previous seven weeks, the average of last week is higher than it. According to the data of the last two weeks, the recent exchange rate has fluctuated more than the previous week.

Figure 4: Description of Euro to US Dollar exchange rate trend

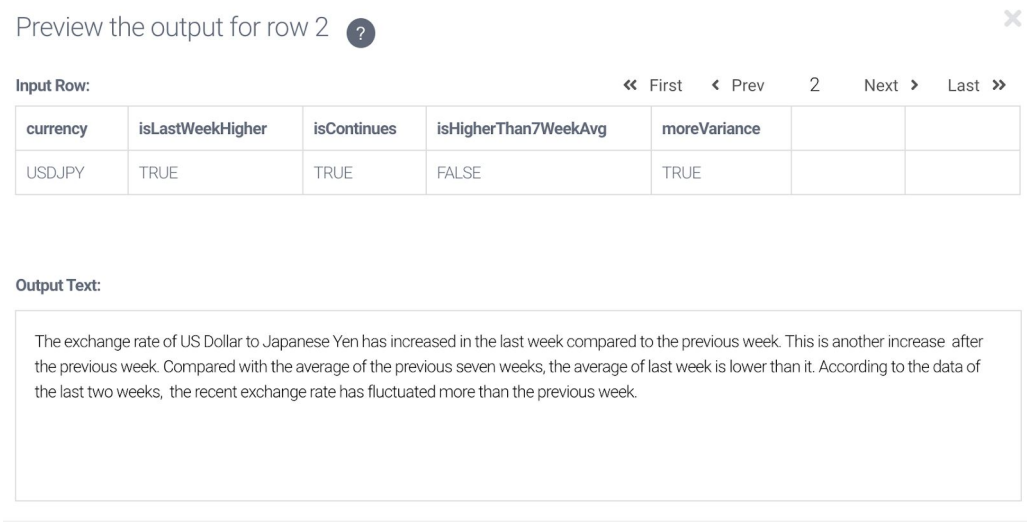


Figure 5: Description of USD to JPY exchange rate trend

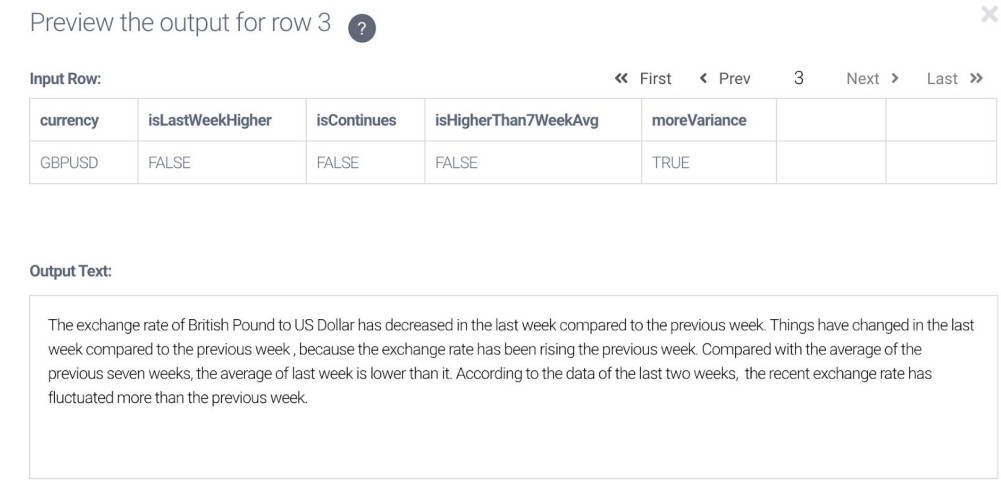


Figure 6: Description of exchange rate trends for British Pound to US Dollar

The chart generated by the data by non-NLG method, because the time span is as long as one year, the line chart is not very beautiful, and users need to summarize the rules according to the generated charts. However, the text generated by the NLG method is clear and intuitive, and users do not need to perform analysis. Through text users can know the recent trend of changes. Compared the two methods, NLG is more convincing than the non-NLG method because the NLG method is a general summary based on calculations based on the given data. However, the non-NLG method only gives charts without any description. Calculations without detailed data are prone to errors.

The second challenge is about vague language. Obviously, the NLG method has been fuzzified, and the relevant detailed data has been fuzzified by the change trend, but the chart expression has not been fuzzified. It is more difficult to understand data through such defuzzified expressions, because people understand fuzzy language more easily and

prefer this form of expression. Based on the comparison of the above two points, the NLG method has more advantages and makes it easier for decision makers to make decisions. Coupled with the voice function, it can be used by more users, so our products are more competitive.

Part 2 – Prototype

Prototype is in *Assignment III - Prototype - README.pdf*

References

Reiter, E. (2019) Natural Language Generation Challenges for Explainable AI.

Deemter, K. van. (2010) Not exactly in praise of vagueness. Oxford;: Oxford University Press.

Gittler, M., 2018. *Quora.* [Online]
Available at: https://www.quora.com/How-does-forex-analysis-website-make-money (Accessed on 15/03/2020).

MatrixTrade, n.d. *Matrix.* [Online]
Available at: http://www.matrixtrade.com (Accessed on 15/03/2020).

Neil, R., 2003. *Currency Overlay.* s.l.:Wiley Finance Series.

Venketas, W., 2019. *DailyFX.* [Online]
Available at:
https://www.dailyfx.com/forex/education/trading_tips/daily_trading_lesson/2019/01/15/forex-market-size-.html. (Accessed on 15/03/2020).

Jhamtani, H. a. G. V. a. H. E. a. N. E., 2017. Shakespearizing Modern Language Using Copy-Enriched Sequence-to-Sequence Models. *EMNLP,* Volume 6, p. 10.

Wei Xu, A. R. W. B. D. R. G.-. m. a. C. C., 2012. *Paraphrasing for style.* s.l., 24th International Conference on Computational Linguistics.

Jhamtani, H., 2017. *Shakespearizing-Modern-English.* [Online]
Available at: https://github.com/harsh19/Shakespearizing-Modern-English. (Accessed on 15/03/2020).

Violeta Gaucan et al. Introduction to the foreign exchange market.
Journal
of Knowledge Management, Economics and information technology, 1(1):1–8, 2010.

Broker Notes. A map of the 9.6 million online forex traders [infographic].
https://brokernotes.co/forex-traders-map/. (Accessed on 15/03/2020).

Zachary M. Seward. This is how much a Bloomberg terminal costs — quartz.
https://qz.com/84961/this-is-how-much-a-bloomberg-terminal-costs/. (Accessed on 15/03/2020).

Internet Usage Stats. Internet world stats - usage and population statistics.
https://www.internetworldstats.com/. (Accessed on 15/03/2020).

Fixer.io. Foreign exchange rates and currency conversion api. https://fixer.io/. (Accessed on 15/03/2020).