

Using Virtual Agents to Bridge the Dutch Digital illiteracy Gap

Derek van den Nieuwenhuijzen
Department of Science
Vrije Universiteit
Amsterdam, Noord-Holland, The Netherlands
d.vanden.nieuwenhuijzen@student.vu.nl

ABSTRACT

The increasing rate at which information is being digitized is putting pressure on low literates in the Dutch society. They are often challenged with the digital environment and more frequently encounter difficulties in completing government forms that are digitized. Although extra budget is spent on numerous projects, there is a lack of best practices and methods in supporting low literates in becoming digitally engaged. New technologies like virtual assistants could potentially empower this group to challenge the digital world. This research is investigating the usage of virtual agents in guiding low literate people with completing digital forms.

KEYWORDS

Digital Divide, Information Science, Digitalization, Social Research

ACM Reference format:

Derek van den Nieuwenhuijzen. February 2019. Using Virtual Agents to Bridge the Dutch Digital illiteracy Gap. Master Thesis research In Proceedings of Vrije Universiteit Amsterdam.

1. Introduction

Since the beginning of the Internet, an increasing amount of data has been transformed from analog to digital. Converting paper-based accounting sheets into digital excel files or converting LP records into digital songs are examples of digitization. Many organizations use digitization to increase their efficiency and to enhance their business. The digitization process caused an extreme increase in the creation of data. Furthermore, for a large portion of people, especially those in developed countries, it caused a high dependency on being connected to the Internet. It has become a necessity in their everyday life to do their work, for educational purposes or to meet with friends. Governments jumped the bandwagon as well and the Dutch government uses a specifically designed strategy 'Nederland Digitaal' to be able to cope with digitization and the increasing threats of cyber-attacks and hackers. The Dutch government website even explicitly mentions to be the number one on the digital forefront in Europe[1].

Despite the ongoing innovations and increasing numbers of internet users, various groups of people lag behind in becoming

digitally engaged. Common terminology for this actuality is digital exclusion, which was already being researched in the beginning of the internet era in 2002 by Servon[2]. In recent years the digital exclusion, or also referred to as the digital divide, gained more attention among researchers and governments. In 2016 the Dutch government asked the 'Sociaal Cultureel Planbureau', a governmental scientific institute, to conduct a large scale research regarding the future of the Dutch citizens. This research highlighted the transformation to a digital society and stressed the importance of having digital skills equal to having social skills[3]. As follow-up, the municipality of Amsterdam conducted a research 'De bouwstenen voor digitale inclusie' (Building blocks for digital inclusion), in which the digital divide of Amsterdam was investigated. The research outlined various digitally excluded groups with the help of interviews and real life examples, and provided specific recommendations based on those use cases[4]. The municipality transformed these recommendations into projects like 'the DigiChallenge' and 'Digimaatjes' (DigiBuddy's) in order to tackle the digital exclusion and to create a complete digitally engaged Amsterdam among all citizens.

One of the most excluded groups that are having difficulties in becoming digitally skilled are low literate people. Research has shown that 1.3 million native Dutch citizens are classified as low literate, of which 300.000 with very poor reading-, writing- or digital skills. Respectively that is 11,9 and 2,6 percent of the entire Dutch population[4]. A high percentage of this group also experiences difficulties in other functional areas such as calculus, problem solving skills and internet usage. Since most information on the Internet is textual information, other than operational skills like handling a computer, one also requires informational skills to assess, and to be able to judge various types of information on their relevance and reliability[5].

Due to the digitization of society, low literate people frequently experience difficulties with digitized forms, online banking, or simply finding an online route description. LostLemon, a research organization in the social domain that develops IT-tools for municipalities, conducted a research among low literate people regarding their perception towards digital forms. The research outlined several barriers that are experienced by low literate people when operating in a digital environment; (i) anxiety in making errors that result in judicial or financial consequences, (ii) forms are experienced as not personal, (iii) webpages contain too

much textual information on a dense area, (iv) user interfaces are complex and crowded with unclear symbols[6]. In collaboration with TNO (The Netherlands Organisation for applied scientific research) LostLemon developed 'Toegankelijke Formulieren' (accessible forms), in which a digitized form was improved to the needs of low literate citizens by using co-design and design guidelines created by TNO. Although the research and the guidelines are helpful in improving the lay-out of digitized forms, and given that research in this domain is yet scarce, there remains a knowledge-gap of other methods in improving these forms for low literates.

This research will focus on bridging the gap in the Dutch Digital Divide by applying new technologies in this domain.

2. Literature review

Digital exclusion has been a research topic as early as the rise of the Internet. Many factors have been investigated to explain the occurrence of digital exclusion in various geographical areas or under specific demographic circumstances. There is an overall consensus among researchers of the social problematics that are accompanied with the digital divide and the ongoing increase of internet diffusion. The main problem is that access to information technologies (IT) is not evenly spread throughout civilizations. This generates a division between the group of people who are benefiting from the considerable benefits of IT and the group of people being excluded, leaving them in a more disadvantaged position in society[7].

The early stages of the digital divide, 1990 to 2005, were mostly defined by access-divide. Having access to a computer directly influences the ability of using the internet and being able to use its benefits. Looking at mostly socio-economic demographic such as gender, age, income and education, Colby classified the digital exclusion in terms of accessibility[8]. As addressed by Rogers, the access-divide evolved into a learning- or content-divide when internet diffusion reached a high level of maturity in many countries by the late 2000's. Especially in developed countries such as The Netherlands, digital exclusion is nowadays mainly caused by the lack of reading-, writing- or digital skills instead of having access to a computer or the internet.

Available research of digital exclusion in relation to low literacy is mainly focused around health issues. Low literate patients frequently encounter difficulties when handling online dossiers or health surveys. Especially eHealth literacy is increasingly attractive in the research domain due to its threatening character on citizens' health. "eHealth literacy is defined as the ability of people to use emerging information and communications technologies to improve or enable health and health care". A study showed that less literate eHealth patients did significantly worse in self-managing their health care. This resulted in a significant higher amount of question asked to a physician relative to a person with a higher eHealth literacy. As a result, less literate eHealth patients felt worse positioned towards a physician than their higher literate counterparts[9].

Research with a different perspective, conducted by TNO, focused on the accessibility of forms by investigating the lay-out of webpages. In collaboration with the municipality of Haarlemmermeer, they developed a new design for an existing webpage on the municipality's website 'Melding Openbare Ruimte'. The design was based on four general design principles that aim to improve user experience: usefulness, usability, desirability and persuasiveness. Furthermore, they incorporated cognitive abilities of users and specifically implemented guidelines for people with a low cognitive ability. Memory, problem solving, focus, reading- writing and understanding language, and calculus were identified by Cremers as important factors in developing an easy-to-use form lay-out. These principles were discussed with a group of low literate people who provided feedback on extra improvements regarding order of sentences and webpages, and usage of visuals. The main adjustments of the municipality's webpage involved the order of webpages, usage of understandable language, use of visuals, avoidance of open questions and being transparent about privacy sensitive information[10].

Other research conducted by TNO in collaboration with Cremers and the Applied University of Utrecht, was concentrated on health questionnaires in order to improve health surveys specifically to the needs of low literate people. By using co-design they developed a new guideline 'Ontwerpen voor laaggeletterden' (designs for low literates) that aims to directly improve the user experience of low literates when using online forms. Using big font sizes, minimal usage of buttons, illustration with pictures and visuals were among the most important recommendations. Figure 1 in the Appendix provides an overview of most applicable guidelines resulted from the research[11].

Most previous research strongly focused on the lay-out to improve webpages for low literate people. However, new technologies, ironically enough, could be helpful in supporting low literates with filling in digital forms. Virtual agents are being used for a variety of applications ranging from improving E-commerce to treatment for depression. These tools mainly use natural language processing, social abilities and learning techniques to offer support in information gathering or other user experiences[12]. Research has shown the visual presence of an agent is critical when looking at motivational and affective outcomes in particular; hence, a voice alone containing the same message is not sufficient [13]. Moreover, a combination of visually present agent with human led voice provides greater learner perceptions and increases the agent's in the context of a learning environment[14].

An initiative by the municipality of Utrecht used a virtual agent to support low digitally skilled citizens with difficult online forms, social service applications or their online banking. By using an interactive and approachable website, Steffie.nl, citizens are guided through application forms by a virtual assistant, Steffie. By providing extra explanation when encountering difficult terms or language and assisted by audio, the virtual agent is able to provide extensive support for low literate citizens[15].

Although Steffie.nl is a great initiative in supporting low literate people, it only provides support about the digital forms through their webpage. Considering a low literate person that is also digitally unskilled, switching from Steffie.nl, to obtain helpful tips, to the digital form could be a challenge in itself. Therefore, this research is focusing on the digital perception of low literate citizens by using virtual agents to improve digital forms. Important to note, the virtual agents are implemented in the digital forms to prevent switching between different webpages. The following research question is the focal point of this research and is to be answered at the end of this Master Thesis.

Research question: “Are virtual agents able to guide low literate people in completing digital tasks?”

A part from the general research question, it was found interesting to investigate the requirements that are necessary when designing a virtual agent for this specific domain. Therefore, this research will investigate the following sub-question.

Sub-question 1: “what are the requirements of a virtual agent?”

Furthermore, being on the Internet is mostly seen as a challenge for low literate people. It is assumed that being digitally skilled improves ones digital experience. It is therefore interesting to research if virtual agents are a helpful in improving the digital perception of low literate people. Accordingly, the following sub-question will be investigated during this research.

Sub-question 2: “What is the digital perception of low literate people when using virtual agents?”

3. Methodology

This research investigates the ability of virtual agents in guiding low literate people to perform digital tasks. Based on the interviews during the context analysis and the literature review, the decision was made to perform a qualitative action research related to the ICT4D 3.0 framework. Action research was chosen due to the practical approach of this thesis. By using an agile methodology, a prototype of a virtual environment is developed that will be tested among low literate people, the target group of this research.

The practical approach of this research is augmented by a qualitative part in order to extract results. Strauss and Corbin defined qualitative research as “any kind of research that produces findings not arrived at by means of statistical procedures or other means of quantification”[16]. Given that this research involves a target group that is difficult to reach, any kind of quantification of the data is highly unlikely due to the limited amount of expected participants. Furthermore, qualitative research is more centered on a naturalistic approach and aims to incorporate context from real life settings[17]. Due to the practical approach used in this thesis, obtaining more in-depth information regarding the prototype

provides better results than doing statistical analysis. Especially since user-experience feedback is invaluable in improving certain methods or prototypes, a qualitative approach will benefit the results of this research the most.

At last, four phases of the ICT4D 3.0 framework are used as a guideline along which this research will use be constructed. The next paragraphs specifies each individual phase and provides detailed insights in how this research will be conducted.

3.1 Context analysis & needs assessment

As part of answering the research question, it was found valuable to gain a more general understanding of the digital divide in the Netherlands from various viewpoints. These different viewpoints were provided by meetings with research consultants from LostLemon, an employee from the municipality of Amsterdam and a social worker from the organization ‘Stichting lezen en schrijven’.

General and recurring findings included the novelty and growing urge of this subject. Despite intensified attention by governments and researchers, it was found that this domain lacks a set of generally accepted best practices and reusable methods. This causes the absence of a unified nationwide approach by the Dutch government to effectively tackle the problem. Although extra budget is given to various municipalities and organizations, projects and initiatives to support the target group are often decentralized without much cohesiveness.

Another recurring observation is the experienced difficulty in reaching low literate people for researching purposes or to test new methods on. Both LostLemon and the municipality expressed certain barriers that keep the target group from being reached in an easy manner. Feelings of shame and the inability to find the right support are addressed as the main causes that low literate citizens are being distant and hard to contact. Although the group of low literate people in the Netherlands is significant, the ‘Toegankelijke Formulieren’ test session conducted by LostLemon was only attended by four participants.

The obtained information from the interviews in combination with the literature review, gave enough perspective to identify the key stakeholders in this challenge. First, the government, and in particular municipalities, are a major stakeholder due to their involvement on both a financial level, acting as sponsors to certain organizations, as well as being directly involved with the target group through various projects. Second, organizations such as LostLemon and ‘Stichting lezen en schrijven’, who are operating and doing research in this domain, are major stakeholders due to their direct involvement and comprehensive knowledge about low literates through practical experience. Notably, their established trust with a small group of low literates gives them a central role in, for example, testing a prototype or newly developed method. At last, low literate citizens are an important stakeholder due to their involvement as end-users of prototypes, methods or projects initiated by the first two stakeholders. Furthermore, they are at the heart of the digital exclusion giving them an strong influence due to their needs.

LostLemon developed a stakeholder map with a central role for the low literates (Appendix figure 2). The map outlines all identified stakeholders and distinguishes between direct and indirect stakeholders, as indicated with a star symbol. Furthermore, the map provides a helpful overview of all concerned organizations that are possible areas to get in direct contact with the target group.

3.3 Requirements

Requirements analysis will take a central role in the design and evaluation of the prototype. A user-centered agile approach is used for the system's development, which is determined by collaborating with users and continuously testing phases [18]. The research consists of the following iterating processes (i) engineering version 1.0 of the prototype by using the obtained information during the context analysis as fundament. Use cases from LostLemon and the 'Ontwerpen voor Laaggeletterden' will be used as guidelines in constructing the prototype's lay-out. Version 1.0 will be tested, by means of a survey, among a group of non-illiterate people who possess a variety of digital skills, (ii) re-engineering of version 1.0 based on the obtained user feedback from the test session. This feedback will be converted into an initial requirements document, which is used to build version 2.0 of the prototype. Version 2.0 will be validated by the industry experts that were interviewed during the context analysis. Their hands on experience with the target group will likely result in helpful feedback that can be used to make improvements on version 2.0. Additionally, their feedback is used to supplement the first draft of the requirements document, (iii) the obtained feedback on version 2.0 will form the basis in constructing prototype version 3.0. This version is to be tested among the target group. The test session is similar to the product testing sessions used by 'Stichting lezen en Schrijven' and LostLemon. The test is conducted by observing the participants in completing an online form using prototype version 3.0. Afterwards, face-to-face interviews with the participants are conducted to obtain user experience information. This feedback will be used to finalize the requirements document and to make adjustments in the version 3.0 prototype.

3.4 Engineering

The prototype will be built in three phases resulting in three different version: version 1.0, version 2.0 and version 3.0. Version 1.0 will serve as the basis upon which the following versions are built by making adjustments. Simplicity regarding lay-out and ease-of-use are taken as the most important factors that will define version 1.0. Furthermore, Steffie.nl is taken as an example due its accessible lay-out and straightforward structure.

RenPy, a visual novel engine, was chosen to build the initial prototype. The functionalities of the tool are capable of meeting all needs that are desired by prototype version 1.0. Furthermore, its accessible and simple lay-out are assumed to be perceived

better relative to other tools. At last, RenPy's programming language is easy to understand and is able to deliver good quality visuals for testing purposes.

4. Timeline

Week #	Activity/Deliverable	Description
8	Detailed planning, Engineering research	Enhance current planning with more detail, Extensive research on tools and create an overview of pros and cons of various options
9	Requirements Document, Arrange test group	More detailed research on requirements and reach out to experts for validation and to make arrangements for testing
10	Start building, Arrange test group	
11	Building, Finalize dates for testing	
12	Building	
13	Finalize version 1.0, Validate version 1.0	Use industry experts to test version 1.0 and receive feedback
14	Revise version 1.0, Finalize version 2.0	
15	Validate version 2.0	Receive feedback from industry experts
16	Test version 2.0, Transcribe results	
17	Test version 2.0, Transcribe results	
18	Process test group feedback, Analyze results	
19	Revise version 2.0, Finalize version 3.0	
20	Validate version 3.0	
21	Test version 3.0, Transcribe results	

22	Test version 3.0, Transcribe results	
23	Process test group feedback, Analyze results	
24	First draft Master Thesis, Feedback from supervisor	
25	Finalize Master Thesis, Defend Thesis (20-06-2019)	
26		

ACKNOWLEDGMENTS

5. REFERENCES

- [1] Rijksoverheid. (2018). Retrieved 17-01-2019 from: <https://www.rijksoverheid.nl/actueel/nieuws/2018/06/16/kabinet-nederland-de-digitale-koploper-van-europa>
- [2] Servon, L. J. (2002). Bridging the digital divide: Technology, community, and public policy.
- [3] De toekomst tegemoet. Sociaal en Cultureel Rapport 2016 (2016). Redactie: A. van den Broek, C. van Campen, J.de Haan, A. Roeters, M. Turkenburg en L. Vermeij. 25-32
- [4] C. Dedding, N. Goedhart, R. Kattouw. (2018). Bouwstenen voor Digitale inclusie.
- [4] Buisman, M., & Houtkoop, W. (2014). Laaggeletterdheid in kaart. Den Bosch: ecbo. 15-19
- [5] Buisman, M., & Houtkoop, W. (2014). Laaggeletterdheid in kaart. Den Bosch: ecbo. 54-55
- [6] LostLemon, Toegankelijke Formulieren 2.0.
- [7] E. Rogers. (2001). The digital divide. convergence. 7(4). 98
- [8] D. Colby. (2001). Conceptualizing the Digital Divide: Closing the Gap by Creating a Postmodern Network that Distributes the Productive Power of Speech. Communication Law & Policy. 6. 123-124.
- [9] E. Neter, E. Brainin. (2012). eHealth literacy: Extending the digital divide to the realm of health information. Journal of medical internet research. 14(1)
- [10] TNO. (2017). Ontwerp formulier Melding Openbare Ruimte. Unpublished TNO report. 1-18
- [11] A. Cremers, M. Welbie, H. Wittink. (2015). Deriving guidelines for designing interactive questionnaires for low-literate persons: development of health assessment questionnaire. Univ Acces Inf Soc.
- [12] Claessen, V. Schmidt, A. Heck, T. (2017) A Study on the Usability and User Perception of Customer Service Systems for E-Commerce. Heinrich-Heine-University Düsseldorf, Germany. 117-119
- [13] Baylor, A. & Ryu, J. (2003). The effects of image and animation in enhancing pedagogical agent persona. *Educ. Comput. Res.* 28, 373–395.
- [14] Baylor, A. (2009). Promoting motivation with virtual agents and avatars: role of visual presence and appearance. *The Royal Society.* 3559–3565
- [15] <https://www.steffie.nl/>
- [16] Strauss, A., & Corbin, J. (1990). Basics of qualitative research: Grounded theory procedures and techniques. Newbury Park, CA: Sage Publications, Inc. 17
- [17] N, Golafshani. (2003). Understanding reliability and validity in qualitative research. The qualitative report, 8(3), 600-601
- [18] B, Losada, M, Urretavizcaya, I, Fernandez-Castro. (2013). A guide to agile development of interactive software with a ‘User Objectives’ –driven methodology. *Science of Computer Programming, Elsevier.* 78. 2268-2270

6. Appendix

Richtlijnen 'Ontwerpen voor laaggeletterden'

Anita Cremers ontwikkelde samen met TNO en de hogeschool Utrecht op basis van een co-design aanpak de richtlijnen voor toegankelijke formulieren. De richtlijnen zijn gebaseerd op een questionnaire om iemands gezondheid te bepalen. Onder de meest toepasbare richtlijnen:

- Groot font
- Illustreren met foto's en icoontjes
- Minimale knoppenset
- Beknopte keuzelijsten
- 'volgende' knop als pijl
- Escape knop
- Één vraag per scherm
- Géén terug knop
- Géén voortgangsindicator
- Introductie en afsluiting waarin doel uitgelegd wordt en aangegeven dat fouten maken mag.
- Instructie video

Anita H.M. Cremers, Marlies Welbie, Kim Kranenborg, Harriët Wittink, Deriving guidelines for designing interactive questionnaires for low-literate persons: development of a health assessment questionnaire, Univ Access Inf Soc 2015

Figure 1: Ontwerpen voor laaggeletterden

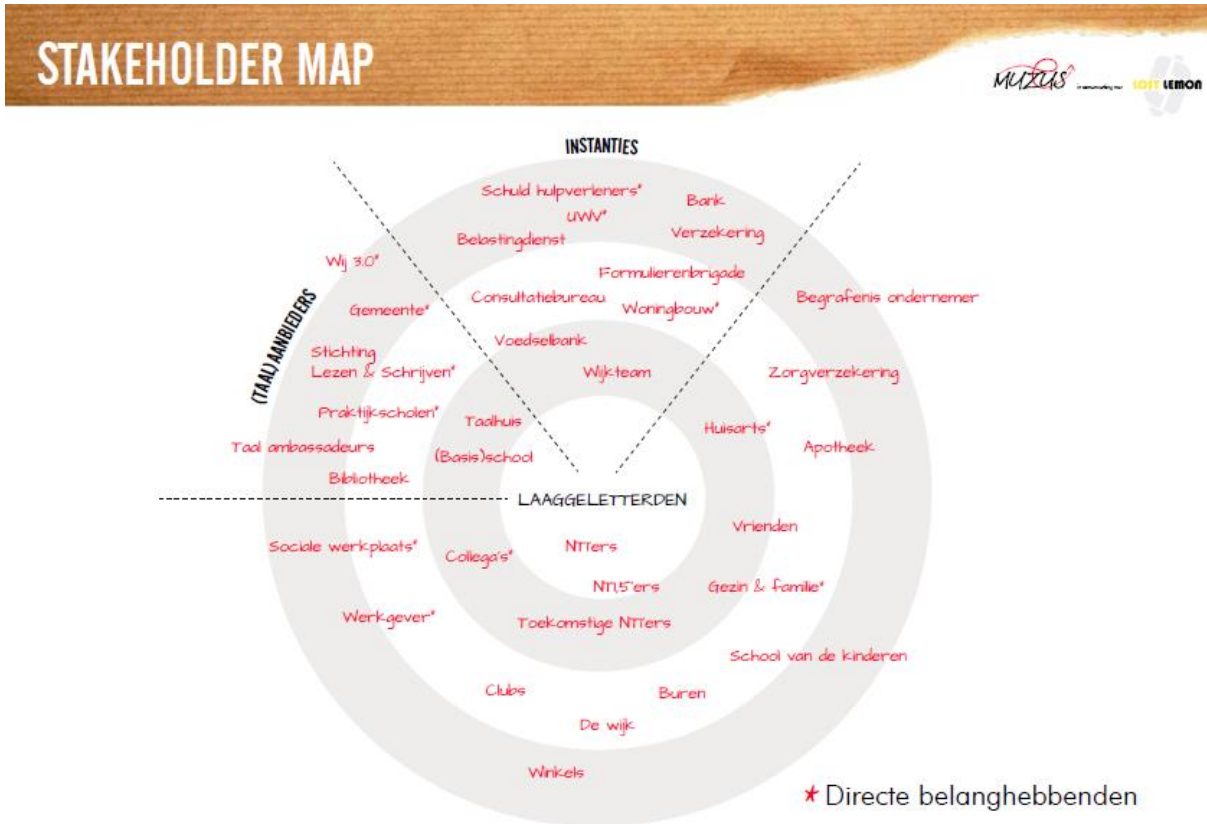


Figure 2: Stakeholder map developed by Lostlemon