

Natural Language Processing Techniques to Reveal Human-Computer Interaction for Development Research Topics

Tawanda Blessing
Chiyangwa[†]

Department of Computer Science
University of South Africa
Johannesburg
chiyatb@unisa.ac.za

Judy van Biljon

Department of Information
Systems
University of South Africa
Johannesburg
vbiljja@unisa.ac.za

Karen Renaud

Department of Computer Science
and Information Systems
University of Strathclyde/
University of South Africa
Glasgow/Johannesburg
karen@strath.ac.uk

ABSTRACT

Designing systems for/with marginalized populations requires innovation and the integration of sophisticated domain knowledge with emergent technologies and trends. Researchers need to be cognizant of existing research trends when aspiring to design interventions to build on current and emergent needs. Traditional manual mechanisms for revealing developments in a field, such as systematic literature reviews (SLRs), cannot meet this challenge because they are time and effort intensive and the domain itself is dynamic and ever expanding. This compromises the efficacy of SLRs in keeping up with the growing academic literature. A number of emergent technologies and modern methods exist that could be harnessed to make it possible to monitor the field more effectively and efficiently. In this paper, we propose the use of natural language processing (NLP), an AI-powered text analysis technique that operates efficiently and requires limited human intervention. To investigate the use and usefulness of NLP for identifying research themes, we applied Latent Dirichlet Allocation (LDA), a topic modelling technique that uses a probabilistic model to find the co-occurrence patterns of terms that correspond to semantic topics. We applied it to a collection of 176 articles published in the Human-Computer Interaction for Development (HCI4D) field. We demonstrate the usefulness of the LDA method by comparing the findings of the LDA analysis to those of a manual analysis carried out by researchers. While NLP techniques may not be able to replace SLRs at this stage, we share some insights on how NLP techniques can complement SLRs to offset investigator bias and save time and effort in revealing emerging domain-related themes.

CCS CONCEPTS

Computing methodologies • Machine Learning • Feature selection

KEYWORDS

HCI4D, Natural language processing, Latent Dirichlet Allocation

ACM Reference format:

Tawanda Blessing Chiyangwa, Judy van Biljon and Karen Renaud. 2021. Natural Language Processing Techniques to Reveal Human-Computer Interaction for Development Research Topics. In 2021 International

Conference on Artificial Intelligence and its Applications (icARTi) *ACM, Mauritius*, 6 pages. <https://doi.org/10.1145/xxx>

1. Introduction

Researchers need to maintain awareness of the latest developments in their domain of expertise, so that they can “build on the shoulders of giants” in their research endeavours. The traditional tool for research topic identification is by using a systematic literature review (SLR), a rigorous and widely-used method of supporting literature analysis in a research domain [23], especially when the aim is to go beyond merely aggregating existing evidence towards constructing evidence-based lessons from the accumulated research. This manual method has the major benefit of being transparent and replicable, but is undeniably time consuming and can be faulted if researchers make contestable decisions (both intentionally and unintentionally) about which works to include and which elements to emphasise in the identified works. There is also a tendency for researchers conducting systematic literature reviews to be inclined towards selecting highly cited papers introduces bias into the process [30].

Machine learning systems, which use computing power to execute algorithms that learn from data [6,8,11,20], might well offer a way to address the potential flaws in SLRs. This triad of computing power, algorithms and big data can combine to facilitate the abstraction of knowledge. When, where, and the extent to which their use is appropriate is still an open question. Moreover, researchers also need to be able to judge when there is a need to lift the lid and peer into the black box, i.e., to revisit some of the properties that were abstracted away during the process [6].

One promising AI technique to harness is NLP, which offers a variety of methods for applying automated topic modelling algorithms to find patterns of term co-occurrence. Such patterns are used to identify coherent topics in text data [1]. One online NLP-based literature analysis tool is iris.ai. It identifies topics and related papers based on a title and abstract, but there is costly and there is little transparency related to the process whereby the resulting topic map emerges.

Considering the domain of interest in this paper, balancing the need for transparency and context sensitivity with the need for efficient and unbiased techniques is critical when deploying AI to make recommendations for marginalized communities [21,29].

Against this background, we pose the question: “How can the principles and practices underlying artificial intelligence (AI) be used to reveal emerging research topics in the developing country context efficiently and effectively”.

The data collection, analysis using the LDA topic modelling method, and findings are presented in the next section. The discussion in the final section compares our findings with those of studies on similar HCI4D data sets and share some insights on how the manual (SLR) and NLP methods can complement each other.

2. Methodological Analysis

We first explain how the corpus was constructed, and then delve into the AI-based analysis.

Gathering the Corpus

As a vehicle for demonstrating the usefulness of the outlined AI techniques we focus on the literature on designing for/with marginalized communities, specifically Human-Computer Interaction for Development (HCI4D). HCI4D is a multi-faceted field that focuses on understanding and designing technologies for under-served, under-resourced, and under-represented populations in diverse geographic regions [3,12,18]. The term “HCI4D” is sometimes contested due to the ambiguity associated with the term “development”. The current discourse favours terminology such as ‘designing for marginalised communities’ or focuses on “Sustainability Development Goals”. However, given the usage history, “HCI4D” was deemed the most appropriate keyword for finding studies related to designing for marginalized communities.

We selected Scopus as one of the recognized Computer Science databases, which indexes ACM articles, where many of the relevant conference proceedings were published. We carried out a search in Scopus to construct an archive of HCI4D research publications published from 2009 to 2019 to support analysis. This selection produced 230 publications, which included the conference proceedings of workshops and symposiums as well as abstracts from books and bulletins published as conference proceedings where the format was condensed and hence not comparable to the other papers. Removing the latter, we were left with 176 publications to analyse. The abstracts of the 176 papers constituted the corpus in this study. Data cleaning, known as text pre-processing, was carried out, removing null values, duplicate data and dealing with outliers. The publications increased from 2009 to 2019 (as can be observed from Fig 1) with a 110% increase in 2013 and a 43% increase in 2018.

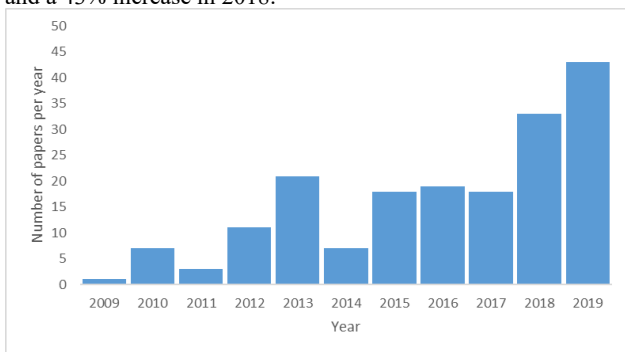


Fig 1. Human computer interaction for development – Publications since 2009 to 2019

Applying NLP

We used the title and abstract of a recent paper about this domain [3] with Iris.ai to provide a comparison with our chosen technique. See Fig 2. Using Iris.ai allows a generic approach since it can be used for any topic but it does not offer the possibility to drill down to constraint the search.



Fig 2: Iris.ai output using title and abstract from [4]

We will now use our proposed AI-powered methodology to demonstrate its greater efficacy in revealing more nuanced insights. Pre-processing clearly shows that the most common words used in the corpus of papers were *design, mobile, paper-based, community, health, use, information, and users* as illustrated in Fig 3.

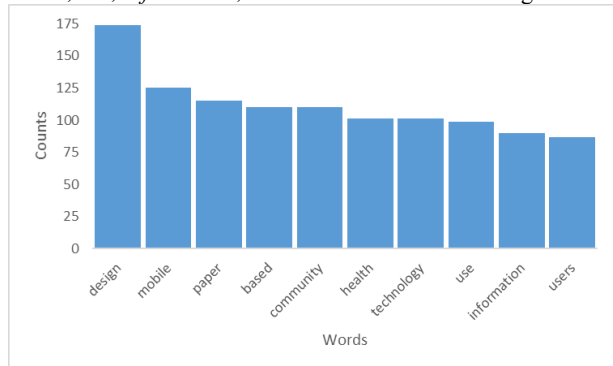


Fig 3. 10 most common words used in corpus papers in this study

More data cleaning steps after tokenization, such as loading text, lexical analysis of the text, stemming and lemmatization was carried out. Jupyter Notebook anaconda software version 3 was used in this study. Hyper-parameter (alpha and beta), number of topics, and optimum iterations were employed as inputs in this study. As a result, α is the magnitude of the Dirichlet prior over a document's topic distribution and β is the Dirichlet prior per-word weight over topic-word distributions [26]. In this study, the value of α is $30/T$ and β is fixed at 0.01, and number of iteration is 30. The letter T stands for the number of topics.

This study used a document-term matrix as a topic modelling technique. Training an LDA model involves finding the optimal set of parameters, under which the probability of generating the training documents is maximized [10]. Ten topics were found to be optimal based on the Topic_Per_Contribution (TPC) loading using a t-distributed stochastic neighbour embedding machine learning algorithm. In this work, we utilized Dodge, Schwartz, Farhadi, Hajishirzi, and Smith [14] to hyper-parameter tune the t-distributed stochastic neighbour embedding machine learning method. Four to ten themes with alpha values of one or five worked well in this study. This research made use of the Natural Language Toolkit and

the genism libraries. The terms generated are depicted in Table 1. Notably, the terms, themes and topic will be used interchangeably in this paper.

Investigating research patterns and trends in software effort estimation, Blei, Ng, and Jordan [5] proposed Latent Dirichlet Allocation (LDA) in 2003 as a generative probabilistic model for collections of discrete data such as text-based corpora. Sehra, Singh, Kaur and Sehra [26] propose the incorporation of topic analysis and labelling to identify latent patterns and trends in datasets and demonstrate a generative statistical method called Latent Dirichlet Allocation (LDA) where the researcher interacts with the results to guide the theme labelling. The documents in a corpus are represented as random mixtures over latent topics with a topic being characterized by a distribution over words [19]. LDA has been selected for our research because it has become one of the most well-known probabilistic models in the field [19] and has been profitably applied to similar problems e.g., analysing research patterns and trends in software effort estimation [26]. Expectation propagation, variation technique, and Gibbs sampling have been utilized by various studies to assess LDA parameters [5,16,22].

Because the study reported here did not create a sample from a joint distribution, Gibbs sampling was not appropriate. To maximize the fit model, the study we report on employed variation Bayes inference [15,25], an extension of expectation maximization. There was no need to produce a sample from a joint distribution, as this technique is supported by expectation maximization.

This was followed by theme labelling; this is where the human experts interact with the results of the algorithm. To mitigate subjectivity and bias two of the three authors did theme labelling individually and then discussed the results. Both suggested the same combination of the themes (see the four groups presented after Table 1), but the naming of the themes was done by consensus to arrive at the four labels suggested namely, HCI4D, D4D, M4D and Context. Based on Table 1, the differences and similarities between the terms are visualized (see Fig 3 in the Appendix).

Nr	Dominant Topic	TPC	Keywords
0	1.0	0.453653	Design, research, researcher, approach, provide, development, development, HCI, local, work, community
1	2.0	0.992180	Community, technology, user, information, design, practice, paper, social, finding, challenge
2	0.0	0.991003	Present, user, mobile, design, interaction, access, study, video, interface, result
3	2.0	0.979440	Community, technology, user, information, design, practice, paper, social, finding, challenge
4	3.0	0.679358	ICT, literacy, technology, network, development, place, practitioner, beneficiary, claim, time
5	3.0	0.744797	ICT, literacy, technology, network, development, place,

			practitioner, beneficiary, claim, time
6	0.0	0.991596	Present, user, mobile, design, interaction, access, video, interface, result
7	3.0	0.603332	ICT, literacy, technology, network, development, place, practitioner, beneficiary, claim, time
8	2.0	0.988965	Community, technology, user, information, design, practice, paper, social, finding, challenge
9	2.0	0.991149	Community, technology, user, information, design, practice, paper, social, finding, challenge

- *Group 0: HCI4D* - the terms cluster around the main theme of designing human-computer interacting for development (HCI4D) which incorporates the user focus and context sensitivity
- *Groups 1,3,8,9: D4D* - the terms cluster around the theme of designing for development (D4D) which incorporates the user and the context in design but does not necessarily focus on the interface or interaction design.
- *Groups 2,6: M4D* - the terms cluster around the theme of designing the interface and the interaction of mobile technologies for development (M4D).
- *Groups 4,5,7: Context* - the terms cluster around the theme of ICT's, Literacy, Practitioners, Beneficiaries, Claims and Time. This seems to resonate with context related constraints like literacy and time. This also highlights the involvement of practitioners and the value to beneficiaries.

Considering the most representative sentence for these four topics provided the results are depicted in Table 2 and those results can be visualized as in Fig 4 (see Appendix).

Topic	TPC	Keywords
0.0	0.991596	design, use, present, mobile, interaction, access, study, video, result, interface
1.0	0.807845	design, research, researcher, approach, provide, development, HCI, local, work, community
2.0	0.992180	Community, technology, user, information, design, practice, paper, social, finding, challenge
3.0	0.744797	Development, technology, network, ICT, literacy, place, practitioner, beneficiary, claim, time

Considering the terms in the Table 2 groups, the keywords of Topic 0.0 (number in leftmost column) with TPC 0.991596 resonates most with the M4D theme. Likewise, Topic 1.0 resonates with the HCI4D theme. Topic 2.0 contains terms from ICT4D in general rather than D4D. Nr 3 is more diffuse, although containing terms which could be linked to the context theme. We acknowledge that there is some subjectivity involved in humans labelling groups but that was mitigated by having the researchers do the labelling independently and then discuss these final suggestions.

Furthermore, this could be managed by a focus group of experts to improve transparency and diminish subjectivity, so the engagement of the experts is less and on a higher level when compared to regular SLR procedure. Finally, the results presented in Table 2 confirmed the relevance of at least three of the four groups which emerged from Table 1.

3. Discussion

Previous systematic literature reviews conducted in the HCI4D field including the studies by Dell and Kumar [12] and Van Biljon [2] applied directional coding analysis [7] based on the most prominent themes (domain questions). Notably, the research themes extracted by the LDA are on a meta-level, as compared to the responses to the domain questions. Therefore, the SLR's themes had to be synthesized based on the responses to the domain question as they could not be extracted directly. For example: a similar systematic literature analyses of the papers published in Scopus from 2007 to 2017 selected on the keyword 'HCI4D' identified three core themes namely, **design**, **development**, and **context** [2]. These core themes included the following sub-themes:

- *design*: design methods, technologies, user participation, usability evaluation methods and ethical practices,
- *development*: sustainability, real-world focus, and in-situ prototyping, and
- *context*: constraints, cross-cultural and unique needs.

The LDA analyses adds value by combining the terms to 'unlock' and render visible latent, previously unseen connections across the textual corpora. By labelling those combinations the researchers could confirm themes like *context* which has been identified in previous literature reviews and become aware of the existence of the M4D theme. Previously, mobile technology has been considered a technology for development [2] but from this analysis, M4D emerged as an independent theme - which did not emerge from the directional analyses.

Our findings confirm a useful synergy between the findings of SLR and the NLP based analyses. Other researchers applied NLP techniques to sentiment analysis and used LDA to automatically extract features of entertainment products based on their descriptions in the form of synopses and summaries from review articles enabling sentiment analysis at finer granularity [1,27]. The findings on review articles' abstract provide evidence that guided LDA can improve the performance of models that predict aggregate outcomes of themes [1,27]. LDA as a topic mixture model is powerful compared to other automated discovery systems which report a simple match or non-match to a group since a fine-grained topical "distance" between two entities can be calculated [17]. From their study with sentiment analysis, we recognized a few areas of potential improvement in our study. Notably, that newer NLP methods strive to analyse segment-level sentiment rather than whole document analysis [17].

This choice of abstracts as the corpus means the results would be expected to align with the core themes which are on a meta level compared to the domain questions of the field. Likewise, the methodology section could be considered to identify the core methodologies or the literature review to identify the core theories, models and framework but using the entire paper may lead to diffuse themes. The findings are limited by the small number of publications but for the sake of comparing the LDA findings with that of the SLR we had to use the same corpus as the publications

selected for the SLR. More research comparing the result from NLP and traditional literature reviews on larger datasets is needed to build trust in the NLP methods and determine the optimal composition of the corpus. Our findings indicate that AI based techniques have untapped potential to support researchers in an efficient and transparent way. The application of LDA has value beyond the HCI4D context but given the importance of context in international development we believe that the applications are particularly useful in exploring this context.

Future research will include themes with extracted short phrases such as bigrams or trigrams instead of limiting it to single words. Since the study is still in its early stages, we considered it prudent to start with a well-known algorithm as a foundation for the research. Moving beyond hierarchical Bayesian models, like LDA we can embed our target documents into some vector space and identify their similarity structure in the vector space by clustering methods.

A range of topic modelling approaches such as Hashtag Latent Dirichlet Allocation (Hashtag-LDA), Enriched LDA (ELDA) and Poisson-Gamma Latent Dirichlet Allocation (PGL-LDA), word2vec, and Bidirectional Encoder Representations from Transformers (BERT) [9,13,19,24,28] should be compared to the basic LDA approach reported on in this paper. Additionally, other databases, such as Institute of Electrical and Electronics Engineers (IEEE) and Web of Science (WoS), may be added to expand the size and diversity of the corpus utilised.

4. Conclusion

The paper reports on an investigation into the use of machine learning tools to increase the efficiency and transparency of topic identification, specifically identifying the core themes from the HCI4D literature. The findings from a systematic literature review of the literature (2009-2019) literature review were compared with the findings of an LDA analysis done on the abstracts of the same corpus of papers. The findings of this study revealed similarities in the themes identified by two different approaches and the possibility of experts engaging with the word clusters in proposing the final themes increase the transparency of the process. Given the efficiency of the LDA analyses and the possibility of reducing researcher bias compared to systematic literature reviews when it comes to uncovering emergent domain-related themes this warrants further investigation towards providing an accessible way for researchers who do not have access to high-end topic modelling tools. LDA was selected as a distinguished tool for latent topic modelling of a large corpus using an unsupervised generative probabilistic technique that extracts themes by selecting the topic with the highest probability. However, future research should focus on more current LDA and vector-based tools.

ACKNOWLEDGMENTS

This paper is based on the research supported by the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation of South Africa (Grant No. 98564). We acknowledge the advice and assistance of Malesela Motlhape and Hulisani Mafune in analysing the data.

REFERENCES

- [1]. Peggy M Andersen, Philip J Hayes, Steven P Weinstein, Alison K Huettner, Linda M Schmandt, and Irene Nirenburg. 1992. Automatic extraction of facts from press releases to generate news stories. In *Third*

- Conference on Applied Natural Language Processing, 170–177.
- [2]. Judy van Biljon. 2020. Knowledge mobilization of human–computer interaction for development research: core issues and domain questions. *Information Technology for Development* 26, 3: 551–576.
- [3]. Judy Van Biljon and Karen Renaud. 2019. Human-Computer Interaction for Development (HCI4D): The Southern African Landscape. In *International Conference on Social Implications of Computers in Developing Countries*, 253–266.
- [4]. Judy Van Biljon and Karen Renaud. 2021. Reviewing a decade of human–computer interaction for development (HCI4D) research, as one of Best’s “grand challenges.”
- [5]. David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research* 3: 993–1022.
- [6]. Alan Borning, Batya Friedman, and Nick Logler. 2020. The ‘invisible’ materiality of information technology. *Communications of the ACM* 63, 6: 57–64.
- [7]. Wichor M Bramer, Gardien B de Jonge, Melissa L Rethlefsen, Frans Mast, and Jos Kleijnen. 2018. A systematic approach to searching: an efficient and complete method to develop literature searches. *Journal of the Medical Library Association: JMLA* 106, 4: 531.
- [8]. Ben Buchanan. 2020. The AI triad and what it means for national security strategy. *Center for Security and Emerging Technology*. <https://cset.georgetown.edu/research/the-ai-triad-and-what-it-means-for-national-security-strategy>.
- [9]. Kenneth Ward Church. 2017. Word2Vec. *Natural Language Engineering* 23, 1: 155–162.
- [10]. Steven P Crain, Ke Zhou, Shuang-Hong Yang, and Hongyuan Zha. 2012. Dimensionality reduction and topic modeling: From latent semantic indexing to latent dirichlet allocation and beyond. In *Mining text data*. Springer, 129–161.
- [11]. Maria De-Arteaga, William Herlands, Daniel B Neill, and Artur Dubrawski. 2018. Machine learning for the developing world. *ACM Transactions on Management Information Systems (TMIS)* 9, 2: 1–14.
- [12]. Nicola Dell and Neha Kumar. 2016. The ins and outs of HCI for development. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, 2220–2232.
- [13]. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- [14]. Jesse Dodge, Gabriel Ilharco, Roy Schwartz, Ali Farhadi, Hannaneh Hajishirzi, and Noah Smith. 2020. Fine-tuning pretrained language models: Weight initializations, data orders, and early stopping. *arXiv preprint arXiv:2002.06305*.
- [15]. Maryam Fatemi, Karl Granström, Lennart Svensson, Francisco J R Ruiz, and Lars Hammarstrand. 2017. Poisson multi-Bernoulli mapping using Gibbs sampling. *IEEE Transactions on Signal Processing* 65, 11: 2814–2827.
- [16]. Thomas L Griffiths and Mark Steyvers. 2004. Finding scientific topics. *Proceedings of the National academy of Sciences* 101, suppl 1: 5228–5235.
- [17]. Alexander Gross and Dhiraj Murthy. 2014. Modeling virtual organizations with Latent Dirichlet Allocation: A case for natural language processing. *Neural networks* 58: 38–49.
- [18]. Azra Ismail and Neha Kumar. 2019. Empowerment on the margins: The online experiences of community health workers. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–15.
- [19]. Hamed Jelodar, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li, and Liang Zhao. 2019. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications* 78, 11: 15169–15211.
- [20]. Yucheng Low, Joseph E Gonzalez, Aapo Kyrola, Danny Bickson, Carlos E Guestrin, and Joseph Hellerstein. 2014. Graphlab: A new framework for parallel machine learning. *arXiv preprint arXiv:1408.2041*.
- [21]. Supreet Mann and Martin Hilbert. 2018. AI4D: Artificial intelligence for development. *Available at SSRN 3197383*.
- [22]. Thomas P Minka and John Lafferty. 2012. Expectation-propagation for the generative aspect model. *arXiv preprint arXiv:1301.0588*.
- [23]. Guy Paré, Marie-Claude Trudel, Mirou Jaana, and Spyros Kitsiou. 2015. Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management* 52, 2: 183–199.
- [24]. Nicole Peinelt, Dong Nguyen, and Maria Liakata. 2020. tBERT: Topic models and BERT joining forces for semantic similarity detection. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 7047–7055.
- [25]. Stephan Reuter, Andreas Danzer, Manuel Stübler, Alexander Scheel, and Karl Granström. 2017. A fast implementation of the labeled multi-Bernoulli filter using Gibbs sampling. In *2017 IEEE Intelligent Vehicles Symposium (IV)*, 765–772.
- [26]. Sumeet Kaur Sehra, Yadwinder Singh Brar, Navdeep Kaur, and Sukhjit Singh Sehra. 2017. Research patterns and trends in software effort estimation. *Information and Software Technology* 91: 1–21.
- [27]. Olivier Toubia, Garud Iyengar, Renée Bunnell, and Alain Lemaire. 2019. Extracting features of entertainment products: A guided latent dirichlet allocation approach informed by the psychology of media consumption. *Journal of Marketing Research* 56, 1: 18–36.
- [28]. Chaojie Wang, Bo Chen, and Mingyuan Zhou. 2018. Multimodal Poisson gamma belief network. In *Thirty-Second AAAI Conference on Artificial Intelligence*.
- [29]. Julie Weber and Kentaro Toyama. 2010. Remembering the Past for Meaningful AI-D. Accessed on 10 August from researchgate.net.
- [30]. Mehmet Yalcinkaya and Vishal Singh. 2015. Patterns and trends in building information modeling (BIM) research: A latent semantic analysis. *Automation in construction* 59: 68–80.

Appendix A

NR	Design	Research	Researcher	Development	HCI	Local	Work	Community	Approach	Provide	Technology	Network	Use	Information	Practice	Paper	Social	Finding	Challenge	Present	Mobile	Interaction	Access	Study	Video	Result	ICT	Literacy	Practitioner	Beneficiary	Place	Claim	Time	Interface	
0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Fig 3. Visualization of topics based on the Topic_Per_Contribution (TPC) loading

Grouping of clusters:

- Group 0: HCI4D - the terms cluster around the main theme of designing human-computer interacting for development (HCI4D) which incorporates the user focus and context sensitivity
- Group 1,3,8,9: D4D - the terms cluster around the theme of designing for development (D4D) which incorporates the user and the context in design but does not necessarily focus on the interface or interaction design.
- Group 2,6: M4D - the terms cluster around the theme of designing the interface and the interaction of mobile technologies for development (M4D).
- Group: 4,5,7: Context - the terms cluster around the theme of ICT’s, literacy, Practitioners, Beneficiaries, Claims and Time. This seems to resonate with context related constraints like literacy and time. This also highlights the involvement of practitioners and the value to beneficiaries.

Identifying themes

NR	Design	Research	Researcher	Development	HCI	Local	Work	Community	Approach	Provide	Technology	Network	Use	Information	Practice	Paper	Social	Finding	Challenge	Present	Mobile	Interaction	Access	Study	Video	Result	ICT	Literacy	Practitioner	Beneficiary	Place	Claim	Time	Interface	
0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Fig 4. Themes based on the Topic_Per_Contribution (TPC) loading of the reduced data set.

- Nr 0 – Themes M4D resonates with the terms design, use, mobile, interaction, access, interface etc.
- Nr 1 – Theme HCI4 resonates with the terms design, research, development, HCI, local, community etc.
- Nr 2 – Theme ICT4D resonates with the terms development, technology, network, ICT, literacy, beneficiary, place, claim, time