SDPS-2021
Printed in US, December 2021

TOWARDS A DATA PROCESSING FRAMEWORK FOR ENHANCED USER CENTRIC SYSTEM ENGINEERING

Fahim A. Salim, Haytham B. Ali and Kristin Falk
University South-Eastern Norway
{ fahim.a.salim@usn.no haytham.ali@usn.no Kristin.Falk@usn.no}

ABSTRACT

There are several challenges faced by organizations in fully utilizing data to enhance their operations. Chief among them are, a lack of common language, ineffective knowledge sharing, information search and ineffective communication among stakeholders. It is our position that designing a data processing framework which generate narrative on demands for users, can help in mitigating some of the challenges. By using industry as laboratory, user centered design and Human Media Interaction (HMI) techniques we aim to utilize the advance in Natural Language Processing (NLP) and other Artificial Intelligence (AI) methods to design and develop and evaluate a framework to enhance the data processing capabilities of the organizations in our test case to allow them to enhance the efficiency of their daily operations. This paper describes over vision of the proposed technological framework.

INTRODUCTION

Organizations generate a large amount of data and information during the project life cycle. However, its utilization in their operations including early phase decision making is quite limited (Falk et al., 2020). The need for "Smarter use of data" is felt vividly by industry (DNV GL AS, 2015). Researchers have observed that they are several challenges which limits the efficient use of big data in decision making (Engen et al., 2021). These among other things include, a lack of common language among engineers and interdisciplinary teams (Tomiyama et al., 2007) ineffective knowledge sharing, difficulty in finding system information (Juzgado P.D. Borches, 2010) and ineffective communication among stakeholders (Delicado et al., 2018) . This paper would use the term data and information interchangeably.

It is our contention that improving information exploration experience of shareholder would improve their utilization of system and operational knowledge. This paper discusses the ongoing research in the H-SEIF 2 project. The H-SEIF 2 project aims to harvest the value of big-data to enhance the experience of stakeholders during complex system engineering project by collaborating with industry partners to improve their digitalization efforts.

While organizations understand the value of data processing in their operations and decision making. However currently the utilization of available information is quite limited. For example, some of the partner organizations in H-SEIF-2 have no processes or methodologies in place to systematically process and utilize the available information, while others to some extents do so. However even in those cases the utilization is limited and there is a lot of room for improvement because:

- The information is scattered around in different systems and is not easily explorable.
- Analysis or data processing is performed on ad-hoc basis, if it all, and communicating the finding or effectively use them in decision making is quite limited.
- There is a lot of knowledge and experience within an organization, which is not being communicated effectively to detect issues/problem early on.

It is the position of this research that advances in the areas of information retrieval, AI driven technologies such as Natural Language Processing, Video/Text/Image/Speech processing, digital engagement & HCI, semantic modeling, and personalization can be used to design and develop methodologies and data processing framework that would help organizations extract more value from their available data sources. Designing such framework would help organizations overcome the above stated challenges by reducing the amount of effort required to explore the information by generating data driven narratives on demand based on the user's information needs.

RELATED WORK

The area of research which deals with the problem of accessing the right information is referred to in the literature as exploratory search. Which is defined as a complex search task in which the user has to first retrieve some facts which then enables further search queries to solve the overall search problem (Marchionini, 2006).

Need for Human Control

Due to evolving user needs, Researchers have long identified the importance of user control in the process of search (Cobârzan et al., 2017). However, the focus has been on creating optimal user interfaces of a predominantly visual character. While these systems do add value to the exploration process, they are quite limited in terms of usage flexibility. System engineering research also stress the need of user control or the importance of user in the loop (Engen et al., 2021; Falk et al., 2020).

Beyond Document Retrieval

Users are constantly exposed to, not only enormous amounts of data, but it is also heterogenous in nature. The standard representation of search results presented by search engines as an ordered list is inadequate for an optimal exploration experience (Jackson et al., 2016).

Therefore while allowing to retrieve needed document(s) from different systems is beneficial for users such as engineers and maintenance personnel (Falk et al., 2020). It is however, tedious and cumbersome for a user to sequence through a long documents to get the desired content (Hong et al., 2011).

To enhance a user's exploration experience with video content, an exploration approach should utilize content-based retrieval and user interaction in a complementary way. It should allow users to interact with content at their own pace, with their own strategies, navigating the content autonomously (Salim, Haider, Luz, et al., 2020).

Generating Narratives by Dashboards and Hypervidoes

Creating a dashboard of visual narratives representation or semantic storytelling is a commonly used technique to provide detail on demand exploration experience to users (Jackson et al., 2016).

Similarly, there are many dashboards style retrieval complex systems. The basic setup of dashboard and retrieval systems is often that they offer access to video summaries of sports activities based on queries to a database which has been filled through manual and automated analysis of what happened in a match or training. Papers may focus on the architectural considerations of sports data retrieval and dashboard systems (Salim, Haider, Postma, et al., 2020) or on the indexing and retrieval itself.

The basic pattern of such systems is: support capture of video and other data, code the data (manually and automatically), augment the video with lots of visual information to make video more informative at a glance, and access the data to derive new insights. To this end, research is also done on novel visualization to give insight-at-a-glance for further exploration (Perin et al., 2018), and on novel forms of querying the large data sets (Shao et al., 2016).

The results need to be communicated, presented to players and others in order to transfer the insights yielded. Most commercial systems support generating rich illustrated reports with data, visualizations, and video materials.

The next step in the process is to support the decision making on top of the generated reports. Many systems are set up to leave this step to the coaches / analysts, but the decision making can also partially be automated (Vales-Alonso et al., 2015).

Hypervideos are a way to consume content in a non-linear manner (Meixner, 2017). They are based on the same notion as hypertext, i.e. hypertext ideas applied on a video content. Following are the primary aspects of all hypervideo based approaches:

- An authoring environment.
- A meta-data structure.
- A specialized environment for consuming content.

Data visualization or dashboard approaches may vary significantly and be hard to understand or require specialized training. The way a user interacts with content is highly dependent on the context of the task (Ganier & de Vries, 2016).In addition to the nature of the exploration task, personal preferences of users in terms of modality of content is also an important factor in providing an optimal exploration experience. While hypervideos give more flexibility to the viewer in consuming the content, the flexibility is still limited to the extent to which the author can anticipate it.

APPROACH

The goal of the research is to use Industry as a Laboratory (Potts, 1993), user centered design (Pinto et al., 2019) and Human Media Interaction (Salim, Haider, Luz, et al., 2020) methodologies to design and propose AI driven approaches to allow users to easily explore relevant information from multimodal data in the context of complex system engineering projects.

First a state-of-the-art review is being performed to identify gaps. Secondly state of affairs at different industry partners is being identified. Design and development of the framework with observe design, prototype and test cycles is performed to:

- Observe personal at industry partners to understand their usage of data or lack of it.
- Design proof of concept of a framework and associated tools, which allows the users to explore the data and extract relevant information efficiently and effectively.
- Evaluate with users and refine.
- Report and disseminate the results.

State of Affairs

To get state of affairs at different partner organizations a survey and semi-structured interviews are being conducted. Industry partners are in the process of providing input to the questionnaire and some personnel would agree to give semi-structured interview. The goal of the survey and subsequent interviews is that partner organization would describe their current processes and systems to process their datasets and documents.

The self-reflection and analysis of the survey and interview would provide a nuance understanding between any gaps and differences at different organizations or at different department within a large organization and a gap between partner organization and the State of Art. This will act is a baseline when making recommendations to the organizations.

Observe, design, prototype, test cycles

The goal is of the research is to design and develop data processing and exploration framework by agile iterative

cycles, using two partners as test case and designing and evaluating the framework and refining it. In the next cycle the framework can be evaluated further with additional partners.

Case Study

As a case study we are focusing on two industry partners. One company is a SME (Small Medium Enterprise) that delivers automated parking garages. The other is a large multinational corporation that provides complete project life cycle services for the energy industry.

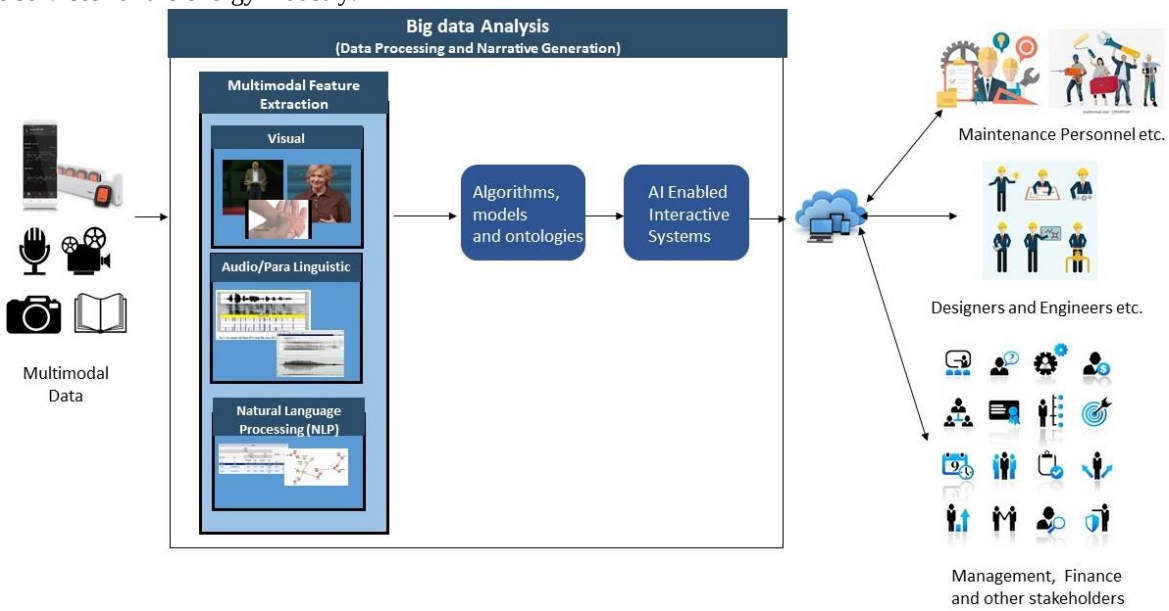

Figure 1: Data Processing Framework (vision).

THE PROPOSED DATA PROCESSING FRAMEWORK

Figure 1: shows the overall visions of using multimodal data and use machine learning techniques to process them and generate a narrative (story from data) according to the user information needs.

While the vision is to be able to use multimodal datasets. In the short run the focus is on Natural Language Processing (NLP) of textual data.

The maintenance and testing teams at both organizations currently use excel sheets to problems occurring in the systems during maintenance and testing. The logs contain manually entered text entries and sometimes they are accompanied with some photographs taken.

Short Term Objectives

In the short term, the goal is to design a framework that it is end to end. It takes input the available data and use AI tools like Natural Language Processing and other Machine Learning toolset to process the data and generate output which would enable the users in this case to see the patterns in the manually entered log e.g.

- What are the most common type of error or problem occurring in their system?
- Are there errors which are only occurring in certain locations and/or certain times?

How the output from the analysis can be used in:

- Communicating between different teams and departments.
- Data driven decision making.

Long Term Objectives

In the long term the aim is design and develop the technical framework (pipeline) to assess how the framework can be utilized in reducing some the challenges organization faced in further utilizing data in enhancing their operations. Following are the research questions we aim to investigate:

- To what extent can the proposed data processing framework can reduce manual effort in creating, updating and exploring the project specific documentation for different stakeholders?
- Can advance AI/NLP techniques aid in better communication among teams and reduce the dissemination of conflicting information?
- How can use the narrative generation add value to the digitalization efforts undergoing at the industry partner?

CONCLUDING REMARKS AND FUTURE DIRECTIONS.

The two organization in our case study are interesting from the research point of view because the difference in their size and operations gives us opportunities in identifying the similarities in the usage of data processing requirement in complex system engineering settings.

The large organization has different digitalization initiatives and utilize many different software systems to

store and explore its data. However, users still face difficulties getting the needed information. While the startup simply does not have any systems in place for data processing. By designing the framework in terms of these case studies we aim to learn how can we improve upon the procedures already in place at the large organization and make sure the startup can learn from them.

ACKNOWLEDGMENT

This research is part of the H-SEIF 2 project and is funded by the Research Council of Norway through the innovation project (IPN project no.: 317862).

REFERENCES

Cobârzan, C., Schoeffmann, K., Bailer, W., Hürst, W., Blažek, A., Lokoč, J., Vrochidis, S., Barthel, K. U., & Rossetto, L. (2017). Interactive video search tools: a detailed analysis of the video browser showdown 2015. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-016-3661-2

Delicado, B. A., Salado, A., & Mompó, R. (2018). Conceptualization of a T-Shaped engineering competency model in collaborative organizational settings: Problem and status in the Spanish aircraft industry. *Systems Engineering*, *21*(6), 534–554. https://doi.org/10.1002/sys.21453

DNV GL AS. (2015). *Industrial Perspective: Digitalization in the Oil and Gas Industry*. https://www.dnv.com/oilgas/download/digitalization-in-oil-and-gas-sector.html

Engen, S., Falk, K., & Muller, G. (2021). The need for systems awareness to support early-phase decision-making—a study from the norwegian energy industry. *Systems*, *9*(3). https://doi.org/10.3390/systems9030047

Falk, K., Kamara, A. K., Brathen, E. P., Helle, K., Moe, P. T., & Kokkula, S. (2020). Digitizing the Maintenance Documentation; A System of Systems in Oil and Gas Industry. *SOSE 2020 - IEEE 15th International Conference of System of Systems Engineering, Proceedings*, 493–499. https://doi.org/10.1109/SoSE50414.2020.9130515

Ganier, F., & de Vries, P. (2016). Are instructions in video format always better than photographs when learning manual techniques? The case of learning how to do sutures. *Learning and Instruction*. https://doi.org/10.1016/j.learninstruc.2016.03.004

Hong, R., Tang, J., Tan, H.-K., Ngo, C.-W., Yan, S., & Chua, T.-S. (2011). Beyond search Event Driven summarization of web videos. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *7*(4), 1–18. https://doi.org/10.1145/2043612.2043613

Jackson, A., Lin, J., Milligan, I., & Ruest, N. (2016). Desiderata for exploratory search interfaces to Web archives in support of scholarly activities. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, *2016-Septe*, 103–106. https://doi.org/10.1145/2910896.2910912

Juzgado P.D. Borches. (2010). *A3 Architecture overviews. A tool for effective communication in product evolution* [University of Twente]. https://doi.org/10.3990/1.9789036531054

Marchionini, G. (2006). From finding to understanding. *Communications of the ACM*, *49*(4), 41–46.

Meixner, B. (2017). Hypervideos and Interactive Multimedia Presentations. *ACM Computing Surveys*, *50*(1), 1–34. https://doi.org/10.1145/3038925

Perin, C., Vuillemot, R., Stolper, C. D., Stasko, J. T., Wood, J., & Carpendale, S. (2018). State of the Art of Sports Data Visualization. *Computer Graphics Forum*, *37*(3), 663–686. https://doi.org/10.1111/cgf.13447

Pinto, J., Falk, K., & Kjørstad, M. (2019). Inclusion of human values in the specification of systems: bridging design and systems engineering. *INCOSE International Symposium*, *29*(1), 284–300. https://doi.org/10.1002/j.2334-5837.2019.00604.x

Potts, C. (1993). Software-engineering research revisited. *IEEE Software*, *10*, 19–28.

Salim, F. A., Haider, F., Luz, S., & Conlan, O. (2020). Automatic transformation of a video using multimodal information for an engaging exploration experience. *Applied Sciences (Switzerland)*, *10*(9). https://doi.org/10.3390/app10093056

Salim, F. A., Haider, F., Postma, D., van Delden, R., Reidsma, D., Luz, S., & van Beijnum, B.-J. (2020). Towards Automatic Modeling of Volleyball Players' Behavior for Analysis, Feedback, and Hybrid Training. *Journal for the Measurement of Physical Behaviour*, *3*(4), 323–330. https://doi.org/10.1123/jmpb.2020-0012

Shao, L., Sacha, D., Neldner, B., Stein, M., & Schreck, T. (2016). Visual-interactive search for soccer trajectories to identify interesting game situations. *IS and T International Symposium on Electronic Imaging Science and Technology*, *0*, 1–10. https://doi.org/10.2352/issn.2470-1173.2016.1.vda-510

Tomiyama, T., D'Amelio, V., Urbanic, J., & Eimaraghy, W. (2007). Complexity of multi-disciplinary design. *CIRP Annals - Manufacturing Technology*, *56*(1), 185–188. https://doi.org/10.1016/j.cirp.2007.05.044

Vales-Alonso, J., Chaves-Dieguez, D., Lopez-Matencio, P., Alcaraz, J. J., Parrado-Garcia, F. J., & Gonzalez-Castano, F. J. (2015). SAETA: A Smart Coaching Assistant for Professional Volleyball Training. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *45*(8), 1138–1150. https://doi.org/10.1109/TSMC.2015.2391258