Project plan

Immersive Automation: automated storytelling, audience engagement and user experience in a news ecosystem

Research partners:

University of Helsinki, Swedish School of Social Science (coordinator)
University of Helsinki, Department of Computer Science
VTT Technical Research Center

Abstract

Publishers are downsizing newsrooms and reducing resources needed for producing engaging content, especially content consumed by smaller communities. Without local customized content there can be no local revenues. At the same time, the input for news, e.g. public and private data, is to an increasing degree published openly in a digital format. However, journalists and editors lack the competence to process this digital data on a large scale. There is a need for a new breed of “meta journalists”, who are able to work with scalable data-driven content. Media companies also collect vast amounts of user data, but lack the skills to combine them with journalistic content and advertisements to create a seamless and engaging user experience. We bring forward a solution to these problems. We suggest an automation of editorial processes, such as the gathering, production, and distribution of content. This allows for new opportunities in the media business, as well as attracts entrepreneurial efforts. Automation also assists conceptual changes in existing media companies. Our aim is to create a roadmap and a demonstration of a future news ecosystem based on automated storytelling, intense audience engagement, and user experience. Stories powered by data and machine learning will lead to a dramatically more personal and customised news experience, with localisation of content as a key feature.

1. Background and motivation

Traditional media companies around the world confront many challenges stemming from the radical digital transformation of the publishing industry. Readers and advertisers have moved away from print towards other platforms over a long time. Publishers face ambidexterity, a need to uphold the old business model while building a bridge to a new operational mode. An essential part of the bridge is automation of editorial processes[1][2][3][4] [5],

[1] Lindén, Carl-Gustav (2017).Decades of Automation in the Newsroom: Why are there still so many jobs in journalism? Digital Journalism, 5:2, 123-140.

[2] Clerwall, Christer (2014). Enter the Robot Journalist, Journalism Practice 8:5, 519-531

[3] Latar, Noam L. (2014). Robot Journalists: ‘Quakebot’ Is Just the BeginningWharton, University of Pennsylvania, http://knowledge.wharton.upenn.edu/article/will-robot-journalists-replace-humanl-ones/ (Retrieved 4.3.2015)

[4] Matsumoto, Rie; Nakayama, Hideki; Harada, Tatsuya; Kuniyoshi, Yasuo (2007) Journalist robot: Robot system making news articles from real world, 1234-1241, Rome, IEEE

[5] Van Dalen, Arjen (2012) The Algorithms Behind the Headlines: how machine-written news redefines the core skills of human journalists, Journalism Practice, 6, 5-6, 648-658.

which will radically transform the way media operates.
Automation implies technology platforms and process automation tools [1] completing any functional activity previously performed manually. The technology in our case consists of newsroom computerisation built on data, for instance software generated news. While the logic of legacy journalism emanates from the idea of creating unique stories to attract a maximum amount of readers and viewers, the computationally generated news are based on massive volumes of texts about similar events or circumstances, and already impact current journalism practices.[2] [3]

We recognize 3 major challenges for the publishing industry and propose a solution for these:

Challenge 1: Publishers are downsizing newsrooms and reducing resources needed for producing engaging content, especially content consumed by smaller communities. Without local customised content there can be no local revenues.

Challenge 2: The input for news, e.g. public and private data, is to an increasing degree published openly in a digital format, but journalists and editors lack the competence to process digital data on a large scale. There is a need for a new breed of “meta journalists”, who are able to work with scalable data-driven content.

Challenge 3. Media companies also collect vast amounts of user data, but lack the skills to combine them with journalistic content and advertisements to create a seamless and engaging user experience Adequate, especially quantitative, methods for measuring user engagement are to a large extent missing.

Solution: An automation of editorial processes, such as the gathering, production, and distribution of content allows for new opportunities in the media business, as well as attracts entrepreneurial efforts. Automation also assists conceptual changes in existing media companies. One example is engaging vast amounts of long-tail mini-audiences consisting of relatives to a junior football player. Automation also allows the media houses to reallocate journalists to more creative tasks that need human authoring.

---

[1] Hoyle Brown, Robert; Roehrig, Paul; Malhotra, Vineet (2015). The Robot and I: How new digital technologies are making smart people and businesses smarter by automating rote work. Teaneck, NJ, Cognizant.
[2] Stavelin, Eirik. (2014). Computational journalism. when journalism meets programming, PhD thesis, University of Bergen.
[3] Karlsen, Joakim, & Stavelin, Eirik, (2014). Computational journalism in Norwegian newsrooms. Journalism Practice, 8(1), 34-48.

2. Objectives, approach and results

2.1 Objectives

The overall objective of IA is:

To create a roadmap and a demonstration, of a future news ecosystem based on automated storytelling, intense audience engagement, and user experience. Stories powered by data and machine learning will lead to a dramatically more personal and customised news experience, with localisation of content as a key feature.

The *more specific* objectives are:

The research consortium aims at responding to the double challenge of the media sector: *to increase revenues from digital venues* and *to produce compelling content cost-effectively*. The consortium will innovate new concepts and services for the media sector, particularly within the field of automatic storytelling and user engagement, especially on smartphones and emerging mobile devices (e.g. wearables, VR-headsets).

Automation and algorithms are often considered autonomous processes, black boxes working on their own with only little human input. However, new technology is always socially constructed, and in this project we look at *balancing the different roles between developers, business people, journalists, advertisers, the public, data producers*, and others who have a legitimate interest in how new information and knowledge is created, distributed, and consumed. By bringing together people from different disciplines and backgrounds we wish to gain insights into the processes that form or reformat our understanding of the world.

At the same time we are looking into how to create and adapt new state-of-the-art technology. We will *develop novel ways of automating news production,* also beyond template-based methods. In particular, we investigate techniques based on text mining and machine learning in order to reduce the manual effort in the production of templates and rules. Another goal is to make the automation algorithms less language-dependent.

From an institutional perspective, the objectives are related to the need for impartial and actionable information to uphold and renew democracy. The demise of newspapers, and other traditional media, that have funded quality journalism, imposes a great risk that citizens will only be served news that are biased and sensationalist. This destroys the trust that forms the base for our society. At the same time, good journalism is the core asset in publishers' strategies, as shown by Financial Times, The Guardian, The New York Times, Helsingin Sanomat, and so many others. Good journalism increases audience engagement, which can then be converted into revenues. The new system will enable journalists to create and distribute standard content

in a more effective way and save expensive resources for tasks that require human creativity and offer high value for the user.

2.2 Our solution

Our solution is based on two central assumptions: The first one is that a close bond between the media and loyal readers should be at the heart of publishers' journalistic and business strategies. Advertisement revenues move elsewhere while reader revenues are growing as people still seem willing to pay for useful and engaging pieces of journalism that they cannot find anywhere else (WAN-IFRA 2016). [4] The second assumption is that news automation will be a central part of solving the problem with shrinking resources and an increased need for engaging and compelling journalism. Immersive automation may sound like a paradox, but we believe in the potential of the radically new combination of "meta journalism", creative computing to produce compelling media content, and personalised distribution.

By immersive we mean "digital technology or images that deeply involve one's senses and may create an altered mental state" (Dictionary.com) or "providing information or stimulation for a number of senses, not only sight and sound" (Collins English Dictionary). Steven P. Dow (2008) [5] notes that immersive and interactive stories have the "potential to trigger a psychological state of embodied narrative engagement (ENE): the combination of the feeling of being in a story world (presence), the feeling of empowerment over unfolding events (agency), and the feeling of being caught up in the plot and characters of a story (dramatic involvement).

The need to radically redefine the media business model is a global concern, and any solutions that are seen as a part of the future field stand a good chance of commercial success.
In this project we look at three specific problems that have led to the demise of traditional newspapers. We propose solutions to these difficulties and lay out a roadmap for advanced computerisation of newsroom processes, which combine automation of content with personalised distribution for smartphones and emerging mobile devices based on user data and geo-localisation.

---

[4] WAN-IFRA, 2016, World Press Trends: Newspapers stress brand strength online; audience revenue share grows further. Available:
http://blog.wan-ifra.org/2016/06/13/world-press-trends-newspapers-stress-brand-strength-online-audience-re venue-share-grows-f [Retrieved 5.7.2016].

[5] Dow, Steven P. Dow, 2008. Understanding User Engagement in Immersive and Interactive Stories. PhD thesis, Georgia Institute of Technology, p. 4.

2.3 Research questions

Research question 1: How to "automate the automation" of news, for instance the creation of news text templates, largely language independently? Existing systems of news automation (Narrative Science, Automated Insights) depend on story templates manually produced by journalists. We aim for state of the art by introducing the automation of news templates.

Research question 2: How to increase media consumers' user engagement by monitoring responses to automatically produced news, and utilise this data in automatic content creation? How to measure user engagement and immersion?

Research question 3: How to create a multi-linguistic technical, organisational, and commercial ecosystem for news based on automated text generation and user experience?

Research question 4: How can news automation serve the public need for fair and unbiased news reporting? How can algorithmic accountability be achieved, and how can journalists evaluate and verify information consisting of digital data?

By combining automation with user data analytics it will also be possible to develop personalised news on a hyper-local level. The world has gone from the anonymous internet to the identified internet. Amazon, iTunes, Facebook, LinkedIn, Google, and Netflix know everything about the user. Media companies are lagging behind in these developments. The Hyperlocal work package in the Next Media project (Kuusisto, 2013)[6] produced promising results that can be utilised here. The design of an ecosystem will depend upon an ethically and legally sustainable utilisation of user data in accordance with the EU's forthcoming General Data Protection Regulation.

The fact that news automation system open up new markets, not only in large scale communication, but will also enable content production for niche audiences that have previously been too expensive to cover in relation to the potential audience size, is of particular interest. Niche content can for example relate to a specific interest, a geographical location, or it can be generated for particular cultural groups in minority languages. A large set of mini-audiences will add up to considerable market volumes.

One main challenge concerns the availability and quality of data. To achieve its goals, the project will focus on verified sources of structured data, that are already available, and can be used with a minimum of standardisation, normalisation, and validation. Some sources, such as data from the Finnish parliament, sports data, weather data, and traffic data, have been identified and analysed for the purpose of turning data into news.

---

[6] Kuusisto, O (ed.), Next Media Hyperlocal uutiskooste 2013. http://virtual.vtt.fi/virtual/nextmedia/uutiskoosteet/Next_Media_hyperlocal_2013.pdf

The other challenge is to write the algorithms that produce news from the data. The project partners will develop automatic/semiautomatic methods to extract templates and other structural information from existing news stories, as well as algorithms that apply relevant templates to render the data as news story.

A key concern is the ability of journalists and editors to adapt into computational thinking in their work, and to learn how to act proactively as well as utilise the structured models in news production and distribution. A new category of journalists is needed, the "meta journalists", who are able to think in terms of processes to create content in scalable and repeatable ways. Traditionally, journalists are used to thinking about news as unique stories about unique events, mostly disregarding the potential for mass production. We believe that besides the generation of final news texts there are also opportunities for other forms of automation, for instance data mining generating news alarms and hybrid forms of man-machine journalism.

2.4 Advantage and novelty

Theoretically this research project is situated within computational journalism, which, according to Young and Hermida[7], refers to "forms of algorithmic, social scientific, and mathematical processes and systems for the production of news", or with a more normative approach "the combination of algorithms, data, and knowledge from the social sciences to supplement the accountability function of journalism" (Hamilton, Turner)[8]. This implies a change in the way "stories are discovered, presented, aggregated, monetised, and archived" (Cohen, Hamilton & Turner[14]). Mass media can be regarded as the infrastructure of freedom of expression and here we are talking about formulas for producing and certifying knowledge, "public relevance algorithms", a key logic governing the flows of information on which we depend (Gillespie[9]).

From a technical point of view, news software robots are computer programs containing algorithms that detail the specific instructions a computer should perform (in a specific order) to generate texts and other content simulating the work processes and end results of journalists. Is there an algorithm for journalism? The answer is yes, but at the moment in a very limited sense. Basic instructions for creating news can be used for guiding computers, and strictly rule-based journalism applied in automated news only suits very narrow domains such as sports and financial earnings reports.

---

[7] Young, Mary Lynn and Hermida, Alfred (2014). From Mr. and Mrs. Outlier To Central Tendencies. Digital Journalism, 2014, 3, 3, 381-397.

[8] Hamilton, James T. and Turner, Fred (2009). Accountability through algorithm: Developing the field of computational journalism, 27-41, Center for Advanced Study in the Behavioral Sciences, Stanford

[14] Cohen, Sarah and Hamilton, James T; Turner, Fred (2011). Computational journalism, Commun ACM, 2011, 54, 10, 66-71.

[9] Gillespie, Tarleton (2014). The Relevance of Algorithms, 167-194, MIT Press, Cambridge, MA

We intend to expand this field in an exponential way by developing tools that use text mining and machine learning to partially automate template production, so that a larger variety of domains, expressions, and languages can be covered in the automated news production.

The interdisciplinary approach of this project allows us to aim at reconceptualising IA to achieve qualitatively new dimensions of user engagement. We approach this innovative leap by exploring techniques from computational creativity to enable generation beyond pre-defined templates, for example digital personal news assistants.

2.5 Effects on media businesses

For the participating media companies the outcome of the project will be threefold. The first part is the creation of new editorial models for lean and scalable content production and distribution, both on a personal level as well as for different customer segments. The second, and equally important outcome, will be the internalisation of innovation methods and development processes in the newspapers. Furthermore, the newspapers will develop and test new business ideas with their commercial partners, and also explore opportunities outside the domain of traditional media business. With a high degree of automation, niche content can be produced at a significantly lower cost, which in turn makes it possible to attract new audiences who are interested in that particular niche. This will generate revenue by being part of new business verticals for the media sector. Furthermore, targeted content and advertising is achievable on a smartphone platform, allowing for reconquering lost advertising revenues and increasing content sales revenue. The service will also improve efficiency by reducing the high costs of human resources in writing and researching articles that are relatively predictable in structure, and instead directing these resources to more creative tasks. In addition to this, the project will create new practical knowledge of how producers of public and private data can cooperate with media companies to achieve knowledge formation and new forms of content production, thus creating opportunities for new business.

There will also be opportunities for new business ventures. The project management team will select a few start-ups which will participate in the project and develop their own offering in accordance with the needs of the participating media companies. One company, Streamr, has already been identified. In the spring of 2016 Streamr received 50,000 Euro in funding together with Alma Media from Google's Digital News Initiative (DNI) to develop financial news automation. The management has accepted the invitation to be part of IA.
For the foundations that support the research, the project will offer a holistic and intriguing approach to research on journalism and media with a clear focus on problem-solving combined with a high potential for unforeseen and exciting outcomes.

2.5.1 Other effects

For the media and journalism researchers, the project will create a unique opportunity to facilitate media innovation, use scientific evidence as input in a focused media development project, and gain new scientific knowledge and research data. The project will also help researchers to evaluate the need for a new orientation of journalism education with regards to future forms of journalistic work.

Beyond these immediate and instrumental effects we apply a democratic perspective on the project by asking how the technological development of IA can enhance and innovate our understanding of mediated communication in a social space. Digital divides have been extensively researched from the point of view of uneven access, competence or motivation among users (i.e., Jan van Dijk, The Deepening Divide: Inequality in the Information Society, SAGE, New York, 2005). This project expands the discussion by focusing on the digital divide in the production and producer/consumer interface. We address the lack of solutions that allow information production, interactive dissemination, and use in small markets, such as small states with their specific social and linguistic market conditions.