Title: Immersive Automation - a new Finnish research project

Author(s): Södergård, Caj

Citation: NxtMedia Conference, 16 November 2016, Trondheim, Norway

Date: 2016

Rights: This article may be downloaded for personal use only.
Immersive Automation
- a new Finnish research project

Prof. Dr. Caj Södergård

Nxt Media conference

Trondheim 16.11.2016
Datorer blir snart journalister

Snart kommer datorer att producera texter för media medan journalisterna själva fungerar som meteredaktörer. Teknologiska forskningscentralen VTT och Helsingfors universitet testar hur duktig en dator kan vara för medias arbete.
Timeline – automatic storytelling

- **2005**: Chicago Crime – Google map mashup
- **2010**: Los Angeles Times. Homecides report: 10 % -> 100 % coverage
- **2013**: Wikipedia – Lsjbot (Swedish)
- **2014**: AP Earning Reports, Automatic Insights
- **2016**: AP Baseball
- **2018**: 20% of business content authored by machines (Gartner)
- **2030**: 90 % of articles written by machines (Narrative Science)
Project Objective

- To create a roadmap and a demonstration of a future news ecosystem based on
  - automated storytelling
  - intense audience engagement

- To study, if stories powered by data will lead to a more personal news experience through localisation of content
The Immersive Automation project

- Volume: 1,1 milj. € (93 person months)

- Duration: 1.11.2016 – 31.5.2018

- 3 research bodies and 9 companies and organisations

- Main funder: Tekes - Finnish Funding Agency for Innovation
  - Within the program "Media Remake"
Participants

- **Research**
  - Univ. of Helsinki, Swedish School of Social Sciences (Dr Gusse Linden, leader)
  - Univ. of Helsinki, Computer Science, (Prof. Hannu Toivonen)
  - VTT (Prof. Caj Södergård)

- **Industry**
  - Media Industry Research Foundation (chair Helene Juhola)
  - Sanoma
  - Alma Media
  - Keskipohjanmaa Kirjapaino
  - Kaleva
  - KSF Media
  - Svenska kulturfonden
  - Conmio
  - Steamr
Work packages

- WP1, Data Specification
- WP2, Automation of News production
- WP3, User Engagement
- WP4, News Ecosystem
- WP5, Skills development
- WP6, Coordination, Dissemination, Exploitation
WP1: Data Specification

Goals:
- Identify and categorise local and other data sources
- Analyse the schedules of new information releases
- Work together with selected journalists

Tasks:
- Identify relevant structured data from different sources
- Use collective creativity (workshops, …) to define potential
- Possibly data produced by the Parliament of Finland
- Explore and define stories that are engaging and relevant from a societal perspective (e.g. role of citizens in a working democracy)
WP2: Automation of News Production

Goal
- Text mining methods to analyse structures, expressions and vocabulary of existing news stories to provide templates

- Algorithms for using such templates to render given data as a new story.

- Design mining and generation as language independently as possible

Tasks:
- Experimental prototype with manual templates in a narrow domain

- An overall architecture for automated template and rule extraction and utilisation

- Text analysis algorithms for automated template production and for their use

- Prototype software for template extraction and utilisation

- Iterative tests of the prototypes and development of the algorithms and prototype
WP3: How does automatic news engage?

- We will study automatically generated content on real audiences
  - what drives user engagement?
  - what creates real immersion?

- Based on the feedback, modify the templates and make new trials.
WP3: How do we measure User Engagement

- **Do an A/B test**: machine / human generated content
  - do the users notice any difference (=Turing test)
  - what level of engagement does arise
  - compare also various machine generated versions

- **Test 1**: interviews and co-creation (N=x00)
  - Web query (e.g. VTT’s Owela)
  - Face 2 face interviews

- **Test 2**: monitor content consumption (N> 200.000)
  - Page views, (e.g. compared to history,…), duration of use, keying frequencies…
  - Social media attraction (clicks, likes, sharing, commenting)

- **Test 3**: measure user affections (N=x0)
  - Eye tracking
  - **Biosignals** measured from audiences (e.g with wearables) to get indications of *emotional arousal*
WP4: News Ecosystems

- Conceptualize and demonstrate a multilingual ecosystem for news
  - Covers technical, organizational and commercial levels
  - Stakeholders: media houses, tech & service providers, data owners,

- The ecosystem is built around automated text generation and proven methods for increasing user engagement.

- The ecosystem contains automated procedures for editorial response to structured data and operates, if possibly, with accessible (FAIR) data

- Create and test standard news templates for scaling up the production of local content in similar contexts

- Explore new ways of distributing content
  - In cooperation with SME companies Conmio and Steamr
  - Includes customization & personalization
Summary: Potential Benefits

- Speed and scale
- Accuracy – but not always (Netflix 2015)
- Make jobs more interesting
- Can produce highly targeted news
- A new type of journalism = Datajournalism 2.0
Summary: Potential pitfalls

- Works still only in limited domains (business reports, sports,…)

- Needs costly training

- Did a human or a machine write the story?

- Cultural resistance

- Machines make mistakes.
TECHNOLOGY FOR BUSINESS