

Titre : Generating narratives from ambient data

Responsable(s) :

Catherine Garbay (Catherine.Garbay@imag.fr) LIG-AMA

François Portet (Francois.Portet@imag.fr) LIG-GETALP

Ehud Reiter (e.reiter@abdn.ac.uk) University of Aberdeen

Mots-clés : Data mining, Data abstraction, Semantic Web, Natural Language Processing, Ambient Intelligence, Artificial Intelligence

Durée du projet : 6 mois

Description

Context and objectives:

This project will investigate the potential for computers to communicate everyday information through automatically generated narratives. Automatic narrative generation is becoming of particular relevance now, as ever larger amount of personal information are collected, often automatically (e.g., location from GPS, medical data from portable sensors, etc) or via social networks (e.g., beliefs or opinions disseminated on the web). Automatic narrative generation will make the communication of such data more human-centric, bringing society closer to meeting the "Big Data" challenge, in which enormous amounts of volatile data need to be efficiently communicated to users while at the same time the development of the Web3.0 will make these data more semantically meaningful and processable.

As a preliminary step towards this objective, the project will aim at building a framework to automatically gather information from the user to summarise it as text in the ski touring or in the hill walking domain. Skiers and walkers make continuous use of on-line portals to share stories about their sorties with their peers. Their stories combine environmental data (weather, snow conditions, slope...), temporal data, and affective/evaluative elements. Users also include images to illustrate some aspects of their sortie that can be used as input data for our purpose. Skiers and walkers share this information both for recreational purposes, but also because it is useful to others who may be planning to make a similar expedition. Sharing stories has the added benefit of forging more cohesion within user communities. Hence, this is a rich domain where we can evaluate the effectiveness of story-generation technology in a real-world context.

Concrete work:

To address this challenge, the candidate will work on the extraction of events from heterogeneous data using approach based on temporal abstraction [2] as well as probabilistic models [1]. This processing stage will results in a graph of events from which the main "scene" of the story will be identified. The main scenes will then be extracted, coherence checked and rearranged according to the main focus of the story. This enriched graph will then be rendered as textual narrative using Natural Language Generation technique [2,3]. The candidate will also conduct human experiment to gather data using wearable sensors such as smart phone in the hill walking or ski domain if possible or in more common everyday life scenario otherwise.

EXPECTED RESULTS

The project is adaptable to the candidates own centre of interest but as example should include at least two of the following items:

1. A state of the art about computational narrative from real data
2. An overall model of the data-to-story generation framework
3. A Software implementation of data abstraction from ambient sensors
4. A Software implementation of a story management system
5. A user evaluation of the implemented framework.

SKILLS:

The candidate is expected to have a good background in object oriented programming (Java or C++). The candidate should express or demonstrate interest in data processing and ambient intelligence as well as Natural Language Processing.

REMARKS:

This internship is expected to lead to a position for a PhD in computing science.

This internship is in collaboration with the Computed Science Department of Aberdeen (Scotland), one of the biggest group in Natural Language Generation

REFERENCES

1. Amate, L., Forbes, F., Fontecave-Jallon, J., Vettier, B., Garbay, C.: Probabilistic model definition for physiological state monitoring. IEEE Statistical Signal Processing Workshop, Nice, France, (2011)
2. Portet, F., Reiter, E., Gatt, A., Hunter, J., Sripada, S., Freer, Y., Sykes, C.: Automatic generation of textual summaries from neonatal intensive care data. Artificial Intelligence 173(7-8) (2009) 789-816
3. Black, R., Waller, A., Turner, R., Reiter, E.: Supporting Personal Narrative for Children with Complex Communication Needs. ACM Transactions on Computer-Human Interaction 19(2) (2012)