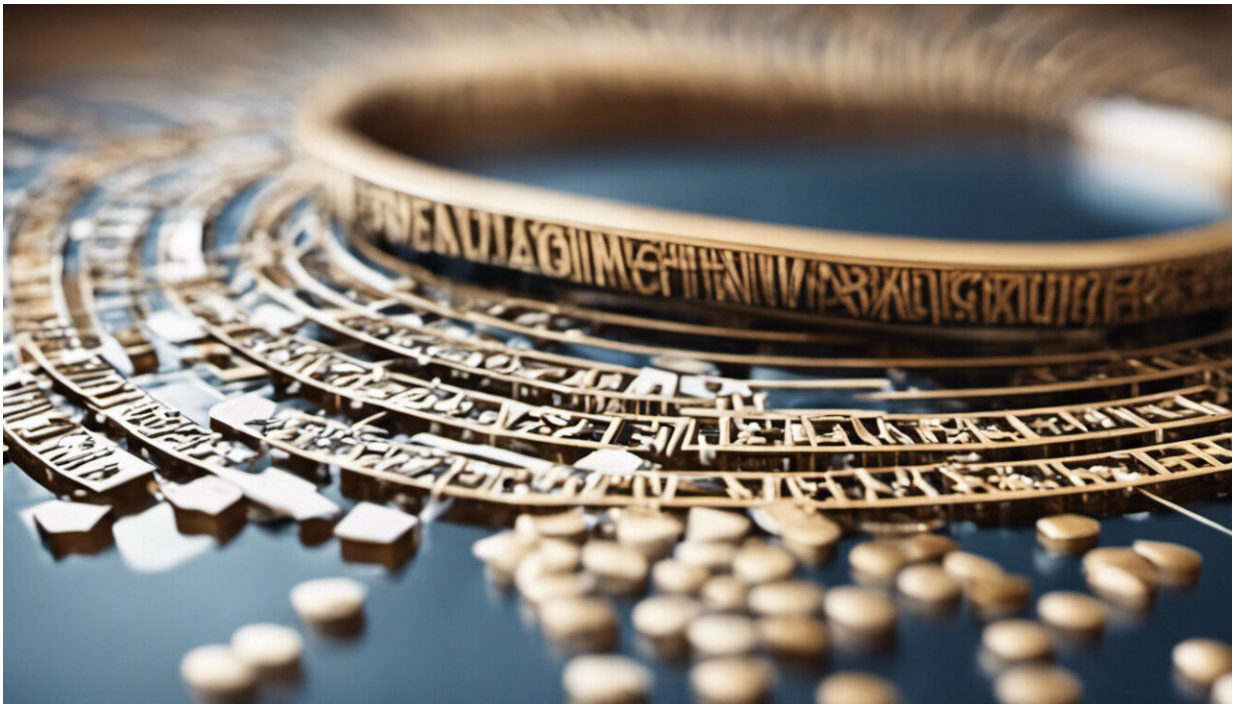# Innovative trust model to help journalists verify social media content

May 19 2016

Credit: AI-generated image (disclaimer)

Using the November 2015 Paris terror attacks as an example, the EU REVEAL project has demonstrated novel solutions for assisting journalists in assessing the accuracy of eyewitness social media content during breaking news incidents.

With a large majority of individuals now actively using platforms such as Facebook or Twitter every day, social media has increasingly become an important source for journalists. In a breaking news event, journalists are now able to pick-up firsthand eyewitness reports that also often contain photos and/or video footage. However, whilst there is lots of genuine information available, it is all too easy for a journalist to accidentally risk their reputation by publishing satire, propaganda or copycat content instead of genuine content during a crisis or emergency news situation.

The REVEAL project has been focusing on developing methods that will allow journalists to quickly and accurately distinguish useful information on social media from 'the noise' – useless or misleading information. They point out that often social media acts as an 'echo chamber', spreading rumours that often turn out to be false. This is not much of a problem for long-term news stories as with time it becomes clear as to what really happened. However, in breaking news situations, it can be much more difficult to quickly distinguish fact from fiction.

Trust model for verifying content

Presenting at the Third Workshop of Social News on the Web in Montreal, Canada, in April 2016, REVEAL project researchers outlined their novel 'trust model' for partially automating the process of filtering useful information on social media by using trusted sources, helping journalists when they need to react quickly to a developing situation. The model allows journalists to maintain a list of their sources, linking new content to authors. When tracking a news story on social media, content items are associated with authors and can be filtered using predefined lists. For each new content item, it becomes clear immediately whether it is in some way related to a source: if it has been posted by that source, mentions that source or is attributed to it.

The model additionally aims to help journalists quickly pick up new eyewitness content. This does not mean trending content from established news organisations or agencies, as the content is no longer breaking. Instead, this would be content that contains eyewitness images or video that is less than five minutes old since publication and is likely to still be unverified.

Paris as an example

To showcase the model and its capabilities, the REVEAL team used the [terror attacks](link) that hit Paris on 13 November 2015 as a case study. Crawling through social media platforms, they used natural language processing techniques to identify named entities (such as 'BBC' and 'Le Monde') in English and French and mentioned URLs. The data was then imported into the trust model, which already contained a sample list of trusted and untrusted sources. By doing this, all content written by, mentioning or attributed to a specific source could be retrieved.

The team then picked five pictures posted during the night of the Paris attacks, with three of them being genuine. They then identified URLs for copies of the posted image that might have been shared instead of the original image URL. They then queried their database in 10 minute intervals during the first hour after each image was published to see how often it was shared (overall and by trusted/untrusted sources). In a second experiment, they sorted URLs by the number of mentions and every five minutes, they compared the currently top ranking URLs that were being shared on social media and filtered the old ones out. By doing this, they tried to detect new eyewitness content to investigate before it went viral.

Analysing the results

When analysing eyewitness content, the team found that untrusted sources generally share images earlier than trusted sources. They also found that trusted sources are an indication for an image to be authentic. Trusted sources that are related to user-generated content make it more likely to be genuine. This is typically the case 30 minutes after a photo has been published. Consequently, if a journalist is prepared to wait, it can point them in the right direction for conventional means of verification, such as factual cross-checking or contacting the source directly through social media channels.

The team also found that for the discovery of newsworthy eyewitness content, it helps to filter old content. Using this method, the 5 tested images showed up in the top 6 % of all content crawled through during a time-window of 5 minutes. This means a journalist does not have to check potentially thousands of social media URLs but can focus on the top URLs.

Although preliminary, these results look promising. The trust model pioneered by REVEAL could help journalists to become both faster and more efficient when sourcing content on breaking stories and publish content with more confidence that material sourced from social media is authentic.

More information: For more information, please see revealproject.eu/

Provided by CORDIS