

Intelligent Analytical News Dashboard

For any broadcaster being the first to broadcast breaking news, monitoring national and international news 7-24 and covering news that is trending and sensitive to its public are of utter most importance. In considering that with today's infrastructure thousands upon thousands of articles are being posted on the web every 24 hours, it is becoming impossibly large for manual, human processing. Today's advances in artificial intelligence provides the media industry the opportunity in analysing big data relating to its assets; in this case its news content as well as its audiences. IRIB's intelligent analytical news dashboard funded by IRIB is working with leaders in data analytics within the academia industry to intelligently analyse news within social media (including websites, social networks and social messaging apps) using the state-of-the-art techniques in AI. The dashboard is designed for journalists, script writers and others involved in news programmes to be informed instantaneously in regard to up-to-date categorized, ranked, summarized and analysed news articles and hence to be encouraged in producing creative content that is important to its readers.



« Micro-service architecture »

The Intelligent Analytical News Dashboard constitutes of six main modules and is presented through a responsive and user-friendly UI/UX:

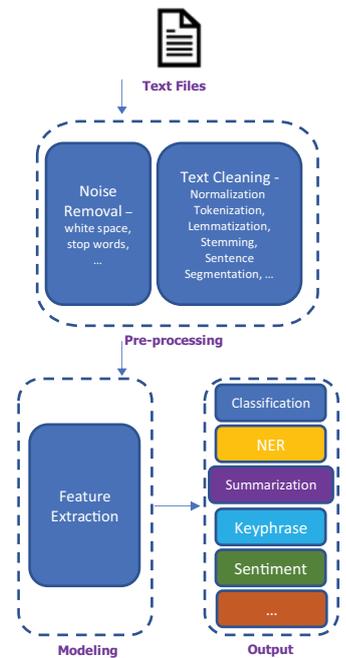
- 1- gathering of news articles from 400 news websites and from hundreds of social media channels,
- 2- storage (sql & nosql),
- 3- pre-processing to unify Farsi text (normalization, tokenization, stemmer, segmentation, ...),
- 4- processing of information,
- 5- visualization tools and
- 6- access to the defined services.

The dashboard will be available to its users through a private cloud and will be designed in a scalable manner. As the system entails text mining modules that can be used and extended within other AI projects, the system is based on a micro-service architecture.

The main services of the dashboard or the micro-services of the system (processing modules) include the intelligent identification of the origin of news, breaking news, news stories and trending news, as well as categorizations of news articles, providing recommendations, extractive news summarization, copy detection, keyword extraction, identification of NER's (named entity recognition - highlighting and tagging jobs, locations, organizations and such alike within articles), sentiment analysis, search and reporting facilities.

« Farsi Text Mining »

The complexity of the dashboard is in the analysis and mining of Farsi textual information. Although within the past decade academic research has been published in this field, due to the complexities of the Persian language and limited available datasets, accuracy results for different modules is a working progress and hence not available in commercial products. The Intelligent Analytical News Dashboard attempts to overcome these challenges using the latest natural language processing and deep learning techniques. The Persian language requires complex pre-processing tasks to provide a unified format to support all its different writing styles that include using or eliminating spaces within or between words, writing words with different spellings, transliterations, Unicode ambiguities and so forth.



Named entity recognition (NER) is used in many NLP applications and classifies and tags named entities that are present in a text into pre-defined categories including “people”, “places”, “organizations”, “professions”, “time”, “date”, “currency” and “events”. Hence, NER can be used to reveal which are the major people, organizations, events and places discussed within input sources (news articles). Knowing the relevant tags for each article will aid in automatically categorizing the articles and enable content discovery.

Topic modeling can be used to automatically classify news articles into trained topics. Furthermore, each news article can be classified into more than one topic e.g. an article can be based both on sports as well as being economically related.

Keywords or **key-phrases** are important components of news articles as they provide a compact representation of the article’s content and are known to be used for search engine optimization. Keyword extraction involves the automatic identification of a sequence of one or more words that best describe the subject of a document.

Automatic extractive text summarization is the process of condensing textual information while preserving the important concepts, which may be based on the combination of statistical, semantic and heuristic methodologies.

Recommendation algorithms can be used to identify and provide suggestions for content that are most likely of interest to a particular user.

Semantic analysis of social media comments can be used to monitor user insights on topics of interest (based on NER – people, events, organizations and etc) and entails the process of defining and categorizing opinions in a given piece of text as positive, negative, or neutral. Search engine to index and search through the incoming data based on specified keywords and to return a list of matching news articles or social media content to the user.

IRIB R&D has made major advances in applying basic research and innovation in media technology. It develops innovative new products and technologies in broadcast/broadband industry that leads the development of future technologies in IRIB. Our R&D projects often involve collaborations with public or private entities, including universities, laboratories, technology start-ups and incubators, research institutes and partner companies.

More Info



Website: rd.irib.ir/en