Article

Extracting Information on Affective Computing Research from Data Analysis of Known Digital Platforms: Research into Emotional Artificial Intelligence

Nafissa Yusupova [1], **Diana Bogdanova** [1], **Nadejda Komendantova** [2] and **Hossein Hassani** [3,*]

1   Department of Computer Science and Robotics, Ufa State Aviation Technical University, 450008 Ufa, Russia; yussupova@ugatu.ac.ru (N.Y.); dianochka7bog@mail.ru (D.B.)
2   International Institute for Advanced Systems Analysis (IIASA), A-2361 Laxenburg, Austria; komendan@iiasa.ac.at
3   Research Institute of Energy Management and Planning, University of Tehran, Tehran 1417466191, Iran
*   Correspondence: hassani.stat@gmail.com

Abstract: The topic of affective computing has been growing rapidly in recent times. In the last five years, the volume of publications in this field has tripled. The question arises which research trends are most in demand today. This can only be judged by analysing the publications that present the results of research. Since researchers have access to the entire global scientific publication space, the task of analysing big data arises. This leads to the problem of identifying the most significant results in the subject area of interest. This paper presents some results of the analysis of semi-structured information from scientific citation databases on the subject of "affective computing".

Keywords: digital platform; affective computing; data analysis; emotional AI

1. Introduction

In the last five years, the volume of publications in this field has tripled (for more details, see [1] for an example). Scientists are frequently faced with questions of relevance of the chosen research topic and analysis of the known results of scientific research in the chosen field. In order to analyse the citation and relevance of scientific publications, openness, and dissemination of scientific results, bibliographic databases of scientific publications have been created. In Russia, this is the Russian Science Citation Index (RSCI) database with the e-LIBRARY information system. This base contains over 12 million scientific studies in Russian. In the international English-speaking community, it is Science Direct. It is one of the world's largest online collections of published scientific research. It belongs to the publisher Elsevier and contains about 10 million articles from more than 2500 journals and more than 6000 e-books, reference books, and scientific collections.

The volume of data contained in such information data systems is quite large and it is practically impossible to analyse all publications relevant to the research topic without automated data processing. The task is to analyse loosely structured information from scientific publications to identify the main research trends in the selected field and to select the most relevant studies to these trends.

In the first stage, a preliminary analysis of publication activity is carried out. Scientific citation databases are analysed by keywords and the most relevant publications on a given topic are highlighted. In the next stage, a semantic analysis of the most relevant publications is carried out.

The Russian Science Citation Index, e-LIBRARY, and the international science citation database Science Direct were selected for the preliminary analysis. The analysis includes the following keywords: "Sentiment analysis", "evaluating emotions", "affective computing", and "tonality analysis". Both scientific citation databases allow:

- Selections to be made based on the phrases of interest in the title of publications, in keywords; and
- Separate analysis by author and organisation, by type of publication, and by publisher.

Quantitative publication metrics can be obtained on request from the databases provided, broken down by year and topic. Analysis of quantitative metrics on the selected topic allows for the identification of significant publications. In the second stage, a semantic analysis of the most relevant publications is carried out to highlight the most interesting scientific findings.

The term "affective computing" was first introduced by Rosalind Picard of MIT, so English-language sources were chosen for analysis to examine the tradition. Interest was aroused by the comparative analysis of scientific trends reflected in English-language and Russian-language publications and the analysis of terms used in this field in the selected languages.

2. Results of Data Analysis in the Science Direct Science Citation Database

The concept of affective computing was introduced by Rosalind Picard of the Massachusetts Institute of Technology (MIT) in 1997. Affective computing is based on recognising and accounting for human emotions in information and cybernetic systems. Emotions in such systems can manifest themselves in different ways, for example, human facial expressions, gestures, voice, and written texts can be analysed. The analysis of emotions is closely related to the analysis of moods and the analysis of the tone of texts. Various emotional signals are used in different applications, and research into this issue is of definite interest. It should be noted that there are different approaches and methods for the classification of emotions, which affects the result of accounting for emotions in different applications. For example, the approach based on the idea of discrete emotions by Tomkins [2]. Another approach to the classification of emotions is based on the assumption of the existence of a multidimensional space of emotions, for example, Plutchik's approach [3]. Different authors use different classifications. The chosen classification must certainly reflect the specifics of the problem being solved and the subject area in which the research is being conducted. The problem of classifications of emotions is deep and requires more detailed analysis and further elaboration in subsequent works. We start with a general overview and try to determine how much research is going on in the broader area of affective computing.

To begin with, let us analyse publications in English from the Science Direct database. Figure 1 present the results of the analysis of quantitative metrics for the selected keywords.

The term affective computing is the most used term in publications. The number of publications with this term is two orders of magnitude higher than with the term sentiment analysis or with the term tonality analysis. In the total number of reviewed publications, 83.9% of the articles include the term affective computing, 0.48% of the articles include the term sentiment analysis, 0.81% of the articles include the term tonality analysis, and 15.11% of the articles include the term evaluating emotions. This may be the fact that sentiment analysis and tonality analysis are particular methods used for evaluating emotions.

Figure 1 shows that before 2010 there was little research interest in the selected topics. From 2010 to 2015, the number of publications increased approximately fivefold. In the last five years, the growth rate of the number of publications has increased significantly. In five years, there have been twice as many publications as before. And in 2020, one-sixth of all publications in the period under review were published. A total of 891,441 publications on the selected topic have been recorded in the database.
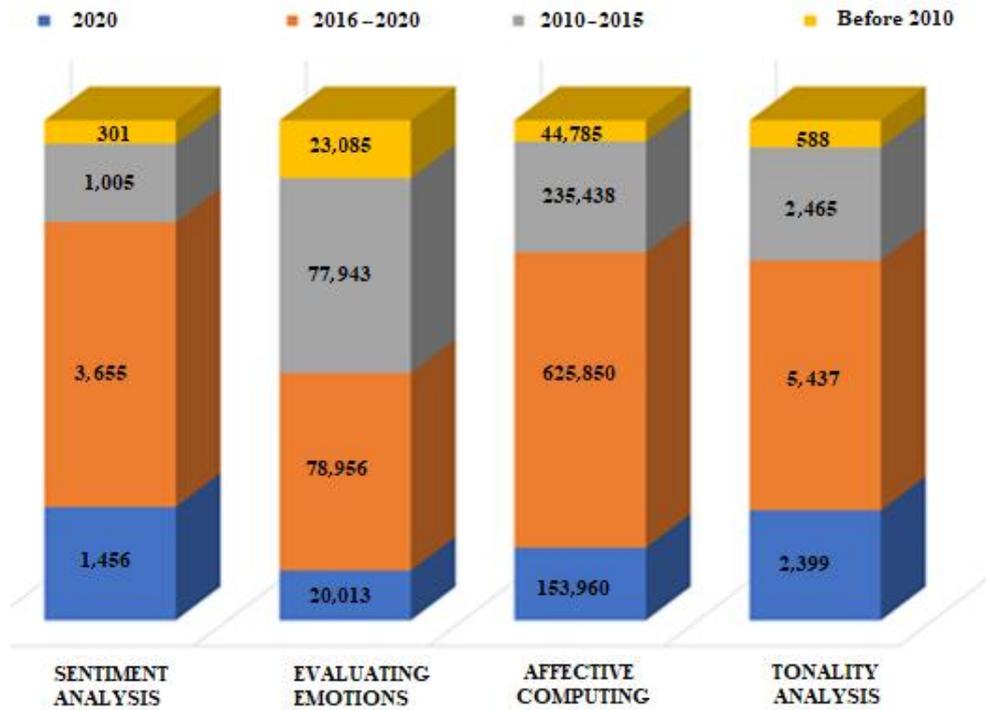
Figure 1. Quantitative metrics of English-language publications, distribution by year of publication.

By comparing the characteristics of the different types of publications (see Figure 2) it can be concluded that most publications are in the form of review articles. This indicates that many scientists are interested in the topic and are studying the accumulated scientific experience.
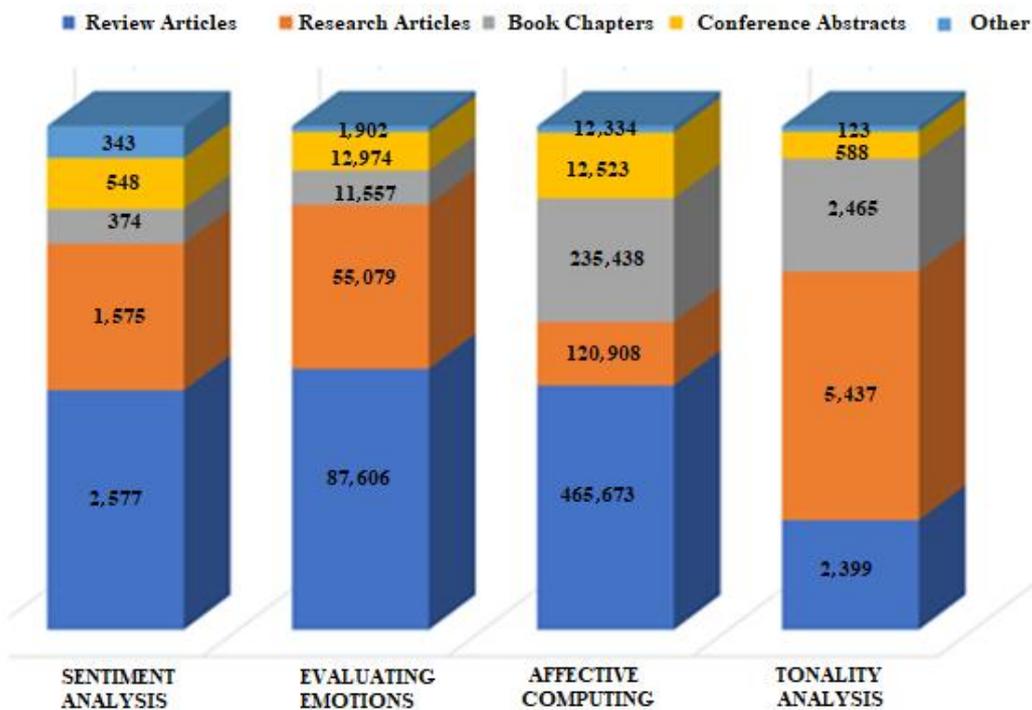

Figure 2. Quantitative metrics of English-language publications, distribution by type of publication.

Scientists have been researching emotions in various sciences for centuries. The development of information technology and artificial intelligence technologies makes it possible to consider the factor of emotions in various applications. The field of affective computing is developing very rapidly, especially in recent times, and more and more scientists and IT developers are interested in its possibilities. This is one of the reasons why review articles are in demand today

The authors identified the main, from their point of view, areas of greatest interest to IT specialists dealing with emotional AI. The Figure 3 shows the developed classification of the key research topics in this area.
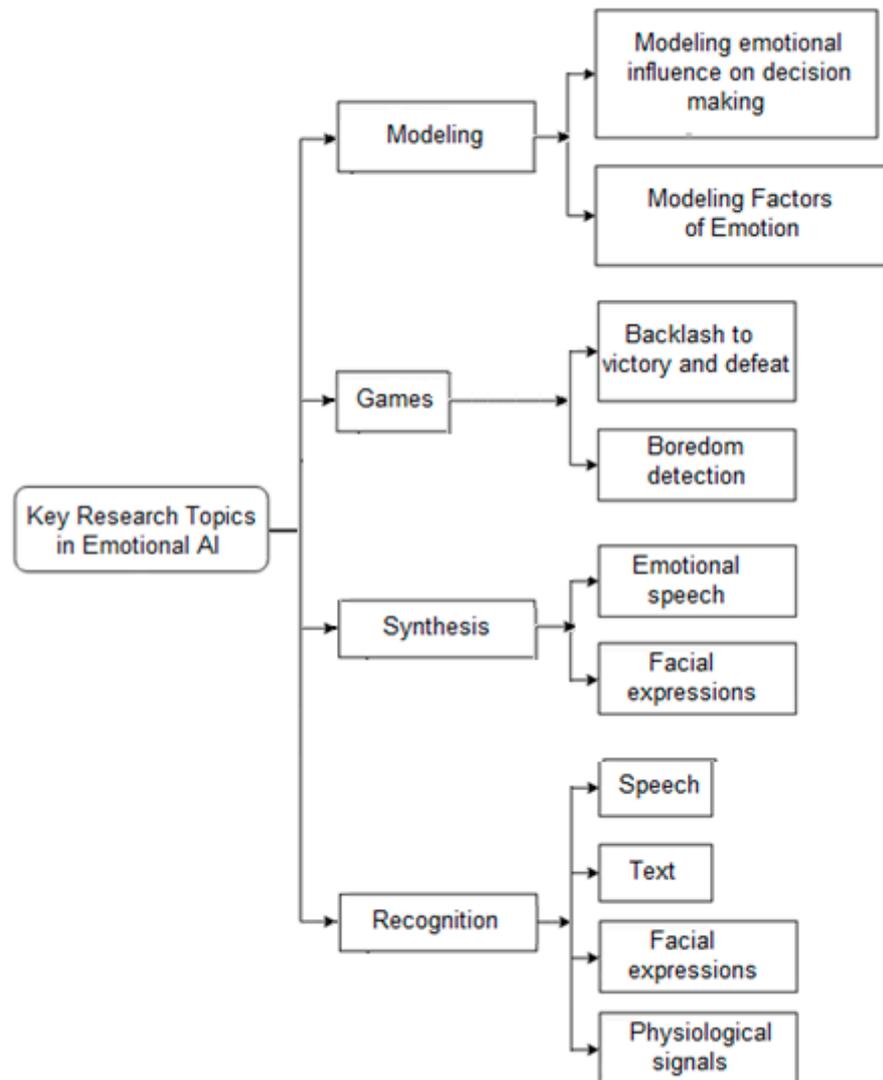

Figure 3. Key research topics in emotional AI.

Figure 4 present quantitative metrics of English-language publications distributed by field of study. The analysis shows that the highest number of publications is observed in the field of technical sciences.
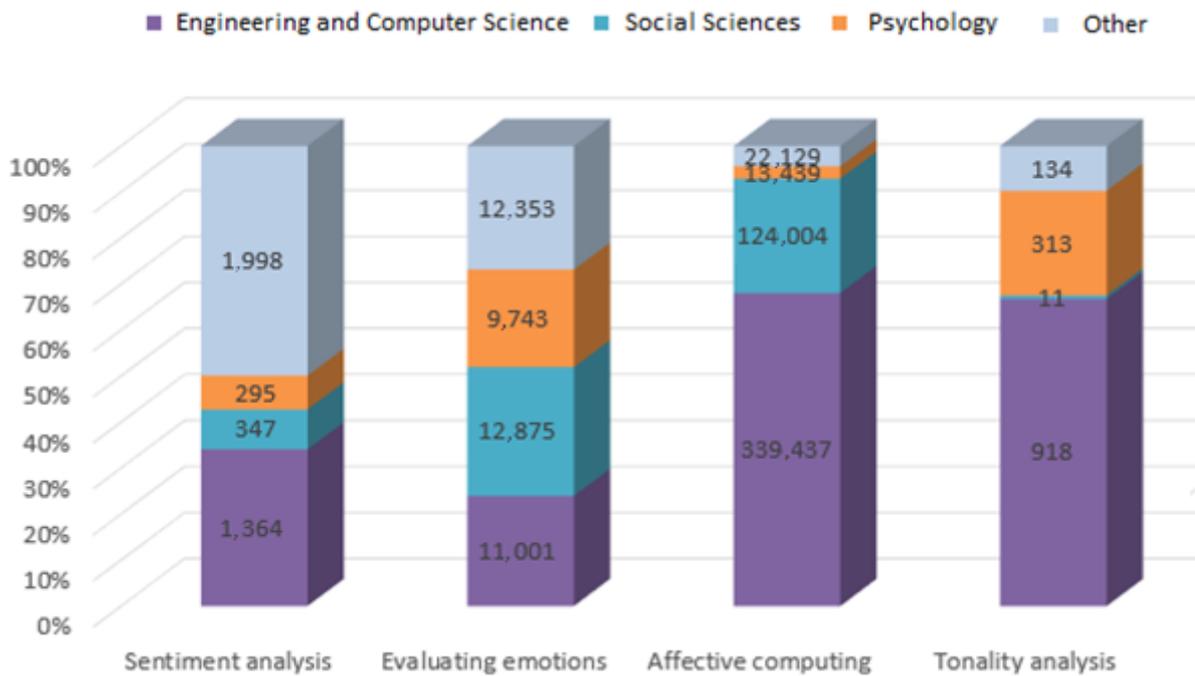
Figure 4. Quantitative metrics of English-language publications, distribution by field of study.

The Science Direct platform provides the ability to analyse publication activity with classification by various subject areas (psychology, medicine and dentistry, neuroscience, social sciences, computer science, business, management and accounting, nursing and health professions, arts and humanities, engineering, etc.).

Emotions and affective computing are found in many subject areas. The authors were interested in research in the field of engineering and computer science, so these areas were consolidated into one group.

An analysis of the dynamics of the number of publications is presented in Figure 5. It can be seen that for all keywords there has been a significant increase in publications over the last 5 years. This indicates the relevance of the research topic and the high interest of the international scientific community.
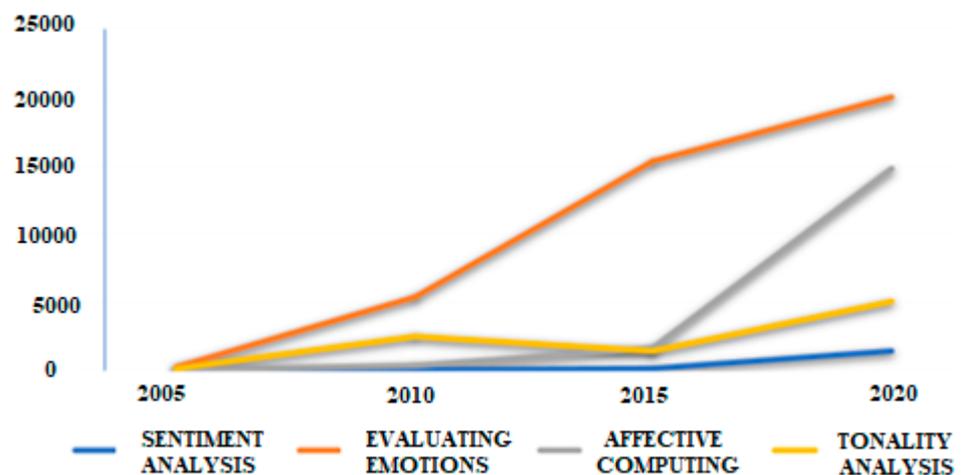

Figure 5. Trends in the number of English-language publications in the selected themes.

We also analysed publication activity by scientific organizations and universities for the keyword "affective computing". The results are presented in Figure 6.
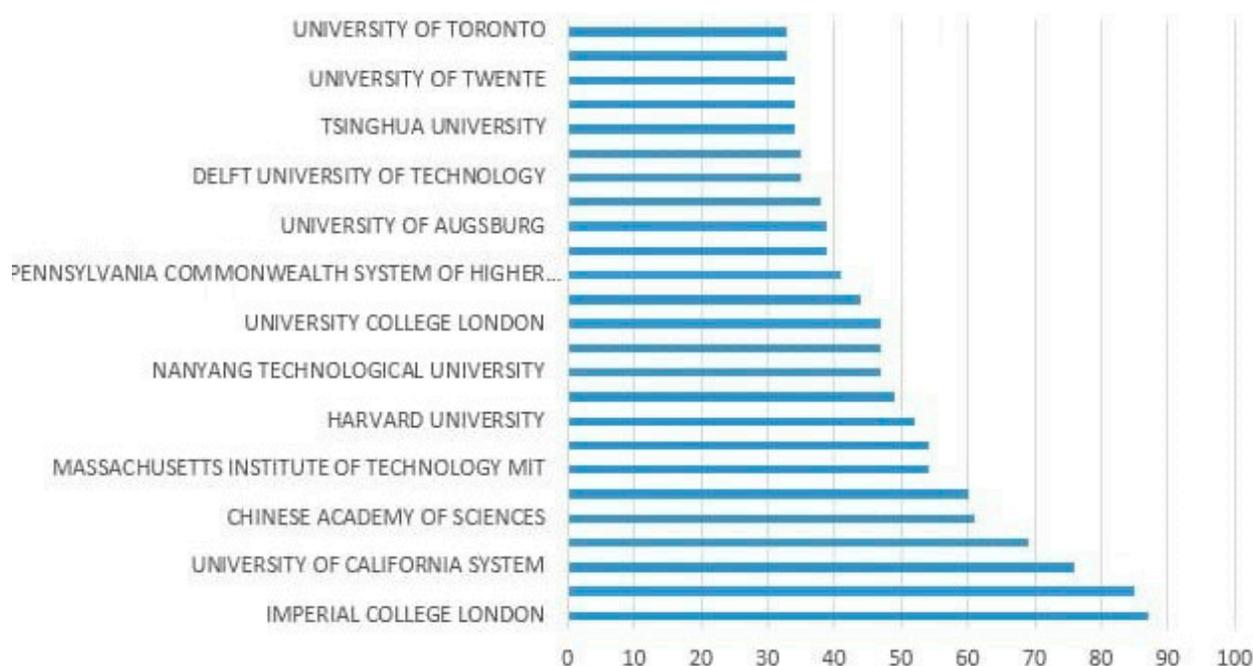
Figure 6. Distribution of the publication activity of research organizations.

The analysis shows that research in the field of affective computing is conducted in different countries. The first place in publication activity for the last 5 years is held by the Imperial College of London (Great Britain), second place by the University of California (USA), and third by the Chinese Academy of Sciences (China). The founder of the term "affective computing", the Massachusetts Institute of Technology (USA), is only in fourth place.

3. Results of an Analysis of the e-LIBRARY Scientific Citation Database

Let us analyse the publications in Russian from the Russian scientific citation database e-LIBRARY. Figure 7 show the quantitative metrics of Russian-language publications distributed by year of publication. Analysing the data, we can say that the concept of sentiment analysis is more in demand among Russian researchers, while the concept of affective computation is just entering circulation among domestic researchers. In general, for the selected keywords, the main activity is in the last five years. The rate of growth of interest in these concepts has increased significantly recently and an overwhelming number of publications were released in 2020.

Publication activity was also analysed by research organisations and universities for the keyword 'affective computing'. The results are presented in Figure 8.

The analysis shows that the number of publications on the selected topics in the individual organisations is not significant. The Voronezh State University and Nizhniy Novgorod State University stand out with the best results.

Figure 9 shows a graph of the dynamics of the number of Russian-language publications on the selected topics. An analysis of the statistics on publication activity in Russian from the e-LIBRARY scientific citation database shows that interest in the selected topics is increasing. The volume of publications for the last five years exceeds the number of publications for the previous period. Researchers in this field are scientific teams of universities and academies of sciences. The volume of foreign research exceeds Russian research by an order of magnitude. Additionally, there is no established terminology in Russian-language literature in the field of affective computation and emotion analysis.
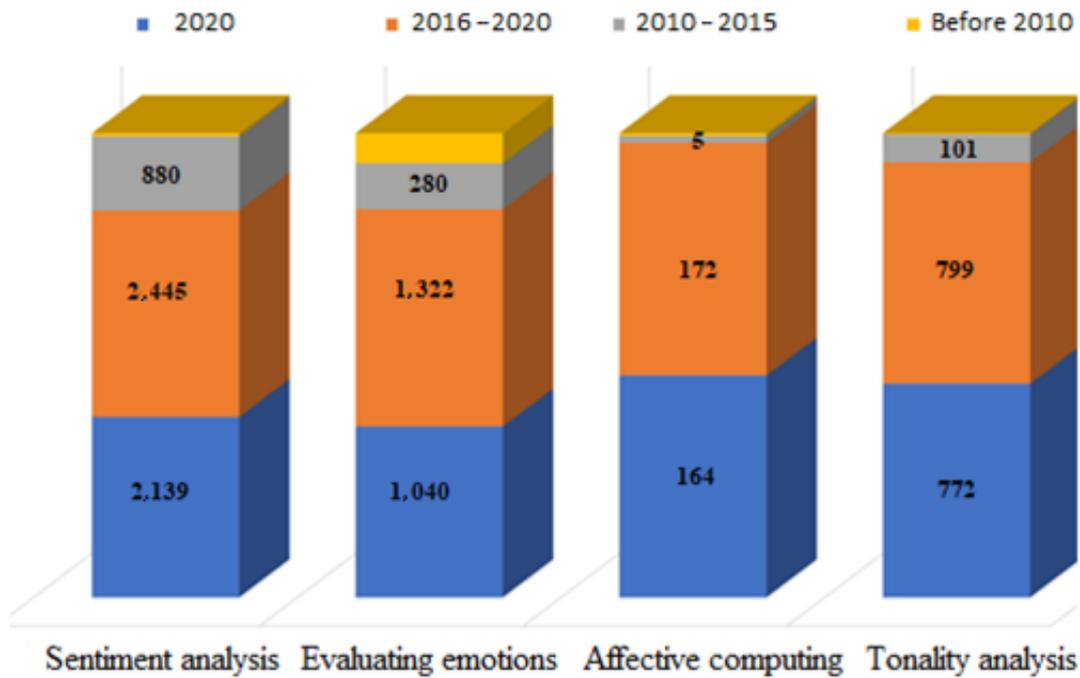
Figure 7. Quantitative metrics of Russian-language publications, distribution by year of publication.
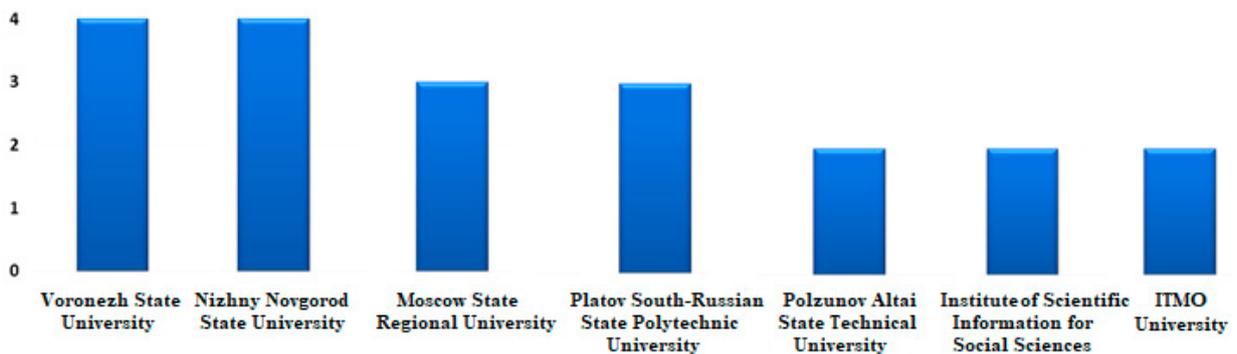

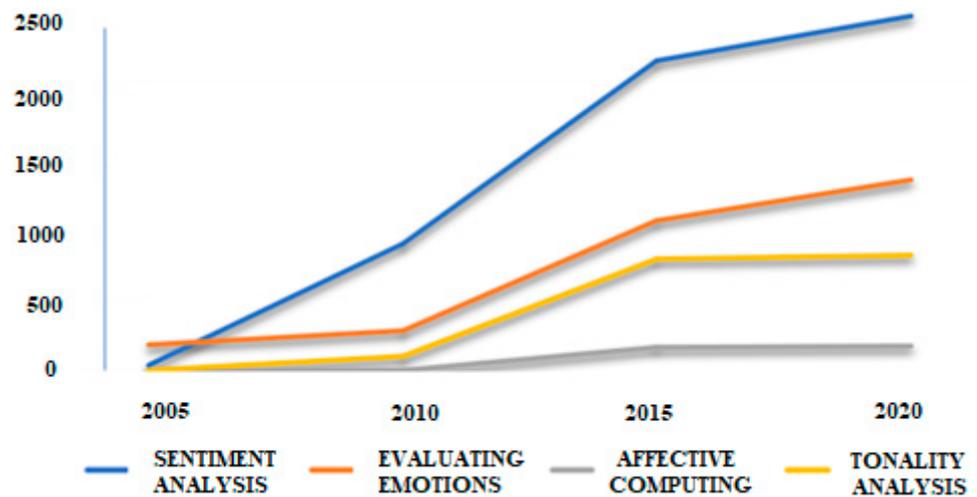**Figure 8.** Distribution of the publication activity of research organisations.

Figure 9. Trends in the number of Russian-language publications in the selected themes.

4. Semantic Analysis after Extraction of Publication Information

Based on the results of the content semantic research and the preliminary processing of the information in the first stage, the most popular publications were identified. A content analysis of these articles allows the most interesting results in this field of science to be highlighted. The main research trends in the subject areas of engineering and computer science are different approaches to emotion recognition taking into account different modalities of emotional information, such as the analysis of texts, voice messages, EEG signals, and video fragments. It is also important to apply emotion information in different applications, especially in decision support in different fields, such as medicine, education, computer games, marketing, etc.

For the semantic analysis, 16 papers reflecting the highlighted scientific trends were selected. Since the bulk of research has been published in the last five years, the analysed list includes works mainly from 2020, 2015, and 2018. Let us consider the main relevant publications in the study area.

In [4], emotion recognition in narrative text is considered in order to adapt decision support to people's emotional states. Machine learning is used to solve this problem by setting up recurrent neural networks with bidirectional processing. A feature of the proposed approach is an adapted form of training. The network is pre-trained for the task of sentiment analysis and the output layer is tuned to the task of emotion recognition. The results of this study show that this approach outperforms traditional machine learning.

The authors [5] in their study analyse vocal media. While conveying semantic information, vocal messages usually also contain abundant emotional information. Data mining is applied to recognise such emotions. In this paper, a computational dynamic method for emotion recognition in vocal social media in the three-dimensional PED (position-arousal-dominance) space is proposed.

Emotion recognition using EEG data analysis is a field of affective computing aimed at finding neural correlates between human emotions and the recorded EEG signals. In [6], the authors cite significant limitations in the application of EEG analysis models for emotion recognition due to the high variability of EEG between individuals. The main challenge in constructing subject-independent affective models is to identify the most distinguishing features between subjects. Analyses were conducted on 26 users, who were exposed to highly arousing affective images. The results are consistent with the neurophysiological hypothesis of a correlation between visually evoked human emotion and brain activity.

Another article [7] provides an overview of research in the field of affective computing. Research on affective computing has increasingly evolved from conventional unimodal analysis to more complex forms of multimodal analysis. Multimodality is defined by the presence of more than one modality or channel, e.g., visual, audio, text, gesture, and eye sensor. The task of fusing information from different modalities is particularly interesting. The authors carried out an extensive study of the different categories of current fusion methods, followed by a critical analysis of potential performance improvements.

The authors of [8] applied wavelet packet analysis to extract emotions from speech signals. The extraction of effective features from speech signals is necessary to recognize different emotions. The results suggest that compared to conventional features, such as Mel-Frequency Cepstral Coefficients (MFCC), EMODB and EESDB features can improve the recognition rate by 14.9% and 4.3%, respectively.

Texts in modern electronic resources have certain specifics, e.g., they can be limited by the length of the text. The work [9] applies machine learning methods and uses three classifiers and three feature selection methods. A special feature of the proposed approach is the use of intra-text features to identify emotions contained in short texts, and the development of a dataset for this purpose. The experiment was conducted on three datasets and six basic emotions were recognized.

In [10], the authors extracted emotions from short video extracts and compared objective and subjective evaluations of emotions. In total 18 video extracts showing staged emotions were analysed. Seven emotions were extracted: neutral, happy, angry, disgust,

fear, sadness, and surprise, which were evaluated using a four-dimensional spherical model of emotions (bad—good; confidence—surprise; attraction; defensive reaction: fear—aggression). Osgood's semantic differential scaling (SD) technique and PXLab-Manikin self-esteem scales were used. The results of the analysis showed that the objective assessment using FaceReader 4.0 and the subjective assessment did not coincide on the same data samples.

The authors of [11] investigated everyday social interaction between two or more people. The focus of the research was on the perception of emotion in terms of event potential and phase synchronisation. The hyper-scanning method for EEG allows the brain activity of a group of people to be analysed. EEG data were analysed using a convolutional neural network to classify categories of emotion. Four emotion categories of anger, disgust, neutral, and happiness were analysed. The results showed that the emotional category of happiness could be classified with higher scores than the other categories.

Another important application of affective computing is in marketing. In the retail market, competition is fierce, forcing us to consider consumers' emotions. In [12], the influence of customers' emotions when receiving a surprise box on the feedback left after purchase was investigated. The analysis showed that consumers write reviews when they experience a vivid emotion. Additionally, there are more reviews when they experience an extremely negative emotion than when they have an extremely positive emotion. In addition, positive and negative emotions play different deterrent roles depending on the product attributes. This research helps firms understand how consumers perceive and value different product and service attributes in the new surprise box business model.

The research described in [13] focused on the analysis of speech emotions. There is a problem of inconsistency in sampling duration and unbalanced sampling categories in the corpus of speech emotions. The authors proposed a model for speech emotion recognition based on Bi-GRU (bidirection gated recurrent unit) and focal loss. The model was improved based on convolutional recurrent neural network training. The experiment was conducted on the IEMOCAP database. The results showed that the speech emotion recognition model proposed in this paper improves the recognition accuracy.

Affective computing has applications in various fields. One such application is the field of education. The work of [14] focused on embedding the user experience of students in student information systems in terms of their emotions, performance, and perceived usability. The study was conducted in a university in Turkey with a sample of 32 participants from different demographic backgrounds. Participants were asked to complete a number of tasks in the system. The tasks were carefully selected to direct the participants towards different aspects of the student information system. In addition, in order to obtain more information about the participants' experiences, they were asked a series of questions. This study showed significant differences between users' emotions, behaviour, and perceived usability depending on their specific demographic background.

An important aspect of affective computing is sentiment analysis. There are many computational approaches in the field of sentiment analysis that have successfully addressed many aspects of this problem, such as classifying data into one of several sentiment categories, etc. The task of integrating and applying these methods to decision making in practice remains relevant. In [15], a general framework for automated sentiment analysis is proposed with a focus on facilitating the model development process for the user in such a way that a good performance can be achieved regardless of the problem domain. The idea behind the proposed approach is to use unstructured opinion-based data in combination with structured data sources for decision making.

The paper [16] describes the analysis of unstructured and textual medical data on infectious diseases. The proposed approach to sentiment analysis uses ontological engineering and analyses aspects by which to measure the views of the general public about infectious diseases. First, an ontology is constructed containing concepts, such as risks, symptoms, methods of transmission, or drugs. The relationship between these concepts is then measured to determine the extent to which one concept influences the others. From

the data obtained, an aspect model of sentiment analysis is constructed, using statistical and linguistic methods. This is done by applying deep learning.

Consumer sentiment was analysed in [17], which is extremely important in the task of understanding customers and developing new products. A vocabulary describing the advantages and disadvantages of products was developed. The authors analysed data from Twitter that is contained in consumer conversations after new product launches in the video game industry. They compared the results of traditional sentiment analysis with those filtered by the lexicon. The results showed a decrease in the number of positive tweets but a dramatic increase in the informative nature of consumers' opinions. Vocabulary-filtered comments offer a much more useful basis for customer insights and new product development.

There are many methods for intelligent analysis of consumer opinions, and most of them can be applied in marketing. The research described in [18] aimed to create an automated way to collect and analyse consumer comments on social media, automatically classifying them into marketing 4Cs and non-marketing categories. Then, they divided the group of marketing 4Cs articles into four types of attribute dimensions to analyse emotional polarity. The merit of this research is the extraction of characteristic keywords from the collected consumer reviews for corpus classification and sentiment polarity analysis, and the creation of a library of functional keywords for specific areas to improve the accuracy of the analysis.

The authors of [19] investigated the sentiments of social media users. The paper addresses the problems of analysing such texts, which are often short, informal, noisy, and rich in linguistic ambiguities, such as polysemy. Another problem is the fact that most existing sentiment analysis methods are based on pure data. The authors proposed a method that applies deep intelligent context embedding to improve the quality of tweets by removing noise while accounting for sentiment keywords, polysemy, syntax, and semantic knowledge and uses a bi-directional network of long-term and short-term memory. The effectiveness of the proposed method was tested on three sets of benchmark data. The proposed methodology for analysing data in digital databases of scientific citations thus makes it possible to highlight existing trends in research.

5. Conclusions

It is not an easy task for a researcher to prove the relevance of the topic chosen for the study. A large volume of publications on the topic of interest must be analysed. The relevance of these publications to the scientific trends in the field under consideration is important in this matter. The approach presented in this paper of analysing large volumes of scientific publications allows the most relevant publications to be selected for further semantic analysis. Statistics of publication activity from the English Science Direct and Russian e-LIBRARY citation databases show an increasing interest in the subject of affective computing and analysis of emotions and moods; the volume of publications in the last five years is many times greater than in the previous period. Further, the vast majority of publications were published in 2020 and most publications are review or scientific articles. The main researchers are research teams of universities and academies of science. The volume of foreign research is an order of magnitude greater than in the previous period. Semantic analysis of the main relevant publications on the subject of "affective computing" showed that the main trends in research are various approaches to emotion recognition, taking into account different modalities of emotional information, for example, the analysis of texts, voice messages, EEG signals, and video fragments. The application of emotion information is important in different applications, especially in decision support in different fields, such as medicine, education, computer games, and marketing.

An analysis of publication activity in English was carried out, since it is the language of communication of scientists from different countries, and native to the founders of the term affective computing. Interest was aroused by the comparative analysis of scientific trends reflected in English-language and Russian-language publications. The next phase of

research plans to expand the data sources and analyse scientific citation databases in other languages.

References

1. Yusupova, N.; Smetanina, O.; Gayanova, M.; Komendantova, N. Semi-structured information in the field of artificial intelligence and information security: Processing results. *IOP Conf. Series Mater. Sci. Eng.* **2021**, *1069*, 012012. [CrossRef]
2. Tomkins, S.S. *Affect Imagery Consciousness: The Positive Affects*; Springer: New York, NY, USA, 1962; Volume 1.
3. Plutchik, R. *A General Psychoevolutionary Theory of Emotion: Theories Emotion*; Elsevier: Amsterdam, The Netherlands, 1980; pp. 3–33.
4. Kratzwald, B.; Ilić, S.; Kraus, M.; Feuerriegel, S.; Prendinger, H. Deep learning for affective computing: Text-based emotion recognition in decision support. *Decis. Support Syst.* **2018**, *115*, 24–35. [CrossRef]
5. Dai, W.; Han, D.; Dai, Y.; Xu, D. Emotion recognition and affective computing on vocal social media. *Inf. Manag.* **2015**, *52*, 777–788. [CrossRef]
6. Bozhkov, L.; Georgieva, P.; Santos, I.; Pereira, A.; Silva, C. EEG-based Subject Independent Affective Computing Models. *Procedia Comput. Sci.* **2015**, *53*, 375–382. [CrossRef]
7. Poria, S.; Cambria, E.; Bajpai, R.; Hussain, A. A review of affective computing: From unimodal analysis to multi-modal fusion. *Inf. Fusion* **2017**, *37*, 98–125. [CrossRef]
8. Wang, K.; Su, G.; Liu, L.; Wang, S. Wavelet packet analysis for speaker-independent emotion recognition. *Neurocomputing* **2020**, *398*, 257–264. [CrossRef]
9. Halim, Z.; Waqar, M.; Tahir, M. A machine learning-based investigation utilizing the in-text features for the identi-fication of dominant emotion in an email. *Knowl.-Based Syst.* **2020**, *208*, 106443. [CrossRef]
10. Vartanov, A.; Ivanov, V.; Vartanova, I. Facial expressions and subjective assessments of emotions. *Cogn. Syst. Res.* **2019**, *59*, 319–328. [CrossRef]
11. Zhu, L.; Su, C.; Zhang, J.; Cui, G.; Cichocki, A.; Zhou, C.; Li, J. EEG-based approach for recognizing human social emotion perception. *Adv. Eng. Inform.* **2020**, *46*, 101191. [CrossRef]
12. Xu, X. Examining the role of emotion in online consumer reviews of various attributes in the surprise box shopping model. *Decis. Support Syst.* **2020**, *136*, 113344. [CrossRef]
13. Zhu, Z.; Dai, W.; Hu, Y.; Li, J. Speech emotion recognition model based on Bi-GRU and Focal Loss. *Pattern Recognit. Lett.* **2020**, *140*, 358–365. [CrossRef]
14. Demirkol, D.; Seneler, C. Evaluation of Student Information System (SIS) In Terms of User Emotion, Performance and Perceived Usability: A Turkish University Case (An Empirical Study). *Procedia Comput. Sci.* **2019**, *158*, 1033–1051. [CrossRef]
15. Kazmaier, J.; van Vuuren, J. A generic framework for sentiment analysis: Leveraging opinion-bearing data to in-form decision making. *Decis. Support Syst.* **2020**, *135*, 113304. [CrossRef]
16. García-Díaz, J.; Cánovas-García, M.; Valencia-García, R. Ontology-driven aspect-based sentiment analysis classifica-tion: An infodemiological case study regarding infectious diseases in Latin America. *Future Gener. Comput. Syst.* **2020**, *112*, 641–657. [CrossRef] [PubMed]
17. Chiarello, F.; Bonaccorsi, A.; Fantoni, G. Technical Sentiment Analysis. Measuring Advantages and Drawbacks of New Products Using Social Media. *Comput. Ind.* **2020**, *123*, 103299. [CrossRef]
18. Lin, H.; Wang, T.; Lin, G.; Cheng, S.; Chen, H.; Huang, Y. Applying sentiment analysis to automatically classify con-sumer comments concerning marketing 4Cs aspects. *Appl. Soft Comput.* **2020**, *97*, 106755. [CrossRef]
19. Naseem, U.; Razzak, I.; Musial, K.; Imran, M. Transformer based Deep Intelligent Contextual Embedding for Twitter sentiment analysis. *Futur. Gener. Comput. Syst.* **2020**, *113*, 58–69. [CrossRef]