Title: Neural Network-based Text Representation for Ranking

In recent years, pre-trained language models have become increasingly popular in natural language processing (NLP) tasks. These models can be used to extract features from text data and preserve the semantic and syntactic information. For example, pre-trained models such as Amharic word2vec, GloVe, fastText, and BERT models, and Amharic BERT tokenizers have been released to reduce the need for heavily-engineered tasks [REF0]. These models can be fine-tuned with labeled data for specific tasks such as text classification, ranking, sentiment analysis, question answering, named entity recognition, natural language inference, co-reference, information extraction, and semantic parsing [REF0].

One example of the use of pre-trained language models is in the development of machine learning comment scoring systems. PerspectiveAPI, developed by Jigsaw, is a machine learning comment scoring system that uses a convolutional neural network to score individual comments according to several attributes using supervised training data labeled by crowd workers [REF3].

In the context of ranking, a fully automatic strategy based on a series of classifiers has been proposed to assess style strength and content preservation. The main classifier uses a pre-trained BERT encoder with a linear classifier on top fine-tuned with a batch size of 256 and sequences truncated at 32 tokens for 6 epochs with learning rate 1e-05 [REF4].

In addition, researchers have explored different variations of architectures to improve the performance of subtasks. For example, the sentence embeddings produced by SBERT are fed to a transformer block containing a multi-headed self-attention layer, its output is then averaged and concatenated with the bag-of-entities representation of the document before being fed to a multi-layer neural network. The output of each layer of this network is also fed to a dedicated neural network that produces the output of each subtask [REF7].

Finally, it is worth noting that pre-processing steps such as normalization, sentence segmentation, and tokenization are crucial for effective text representation. For example, in Amharic language processing, normalization of certain orders and tokenization using WordPiece tokenizer have been shown to be effective [REF8].

In conclusion, pre-trained language models and automatic strategies based on classifiers have shown promising results in ranking tasks. Effective pre-processing steps are also crucial for effective text representation.

References given to GPT:

[REF0] - paperID: 7b8fe8c28a371120b4479540b2c8a0f7c5af25bf Title: Learned Text Representation for Amharic Information Retrieval and Natural Language Processing Chunk of text: We trained and released our pre-trained Amharic word2vec, GloVe, fastText, and BERT models, and Amharic BERT tokenizers, which can motivate the research community to carry out experiments on the pre-trained Amharic language models. Pre-training is fairly expensive, but it is a one-time procedure. Our pre-trained models can be used to reduce the need for many heavilyengineered tasks. The extracted features can be fed into a machine learning model so as to work with text data and preserve the semantic and syntactic information. The constructed Amharic models can be used in different applications. For example, our word embedding vectors can be used for word analogy tasks, namely entity recognition and chunking. The pre-trained Amharic BERT models can also be fine-tuned with some labeled data for a range of specific tasks, such as text classification, ranking, sentiment analysis, question answering, named entity recognition, natural language inference, co-reference, information extraction, semantic parsing, etc. 6.

[REF1] - paperID: 336e531a59cafbe215b950fd749bca866b89cea0 Title: SNK @ DANKMEMES: Leveraging Pretrained Embeddings for Multimodal Meme Detection (short paper) Chunk of text: <https://cloudblogs.microsoft.com/open>source/2020/01/21/microsoft-onnx-open-sourceoptimizations-transformer-inference-gpu-cpu/ Polignano, M., Basile, P., de Gemmis, M., Semeraro, G., & Basile, V. (2019). Alberto: Italian bert language understanding model for nlp challenging tasks based on tweets. In Proceedings of the Sixth Italian Conference on Computational Linguistics (CLiC-it 2019). CEUR-WS.org. Petroni, F., Rocktäschel, T., Lewis, P., Bakhtin, A., Wu, Y., Miller, A. H., & Riedel, S. (2019, September 04). Language Models as Knowledge Bases? <https://arxiv.org/abs/1909.01066> Pushp, P. K., & Srivastava, M. M. (2017, December 23).

[REF2] - paperID: 336e531a59cafbe215b950fd749bca866b89cea0 Title: SNK @ DANKMEMES: Leveraging Pretrained Embeddings for Multimodal Meme Detection (short paper) Chunk of text: D ANDI 0.833 0.834 15.409 NON-CAPISCO 0.785 0.787 35.663 KONKRETIKA 3 0.663 0.668 28.613 KONKRETIKA 1 0.651 0.667 29.933 Baseline 2 0.554 0.567 38.451 KONKRETIKA 4 0.542 0.545 29.836 CAPISCO CENTR 0.542 0.538 48.864 KONKRETIKA 2 0.541 0.545 30.322 CAPISCO TRANS 0.504 0.501 29.927 Baseline 1 0.382 0.377 31.738 withdrawn run3 -0.013 0.067 41.109 withdrawn run1 -0.124 -0.123 44.068 withdrawn run2 -0.127 -0.129 43.890 Table 3: Results for each run on Italian test set. System run Spear Pears Eucl. D ANDI 0.749 0.749 19.950

[REF3] - paperID: 336e531a59cafbe215b950fd749bca866b89cea0 Title: SNK @ DANKMEMES: Leveraging Pretrained Embeddings for Multimodal Meme Detection (short paper) Chunk of text: The AMI task includes both raw (natural Twitter) and synthetic (template-generated) datasets. The raw data consists of Italian tweets manually labelled and balanced according to misogyny and aggressiveness labels, while the synthetic data is labelled only for misogyny and is intended to measure the presence of unintended bias (Elisabetta Fersini, 2020). 2 Background Jigsaw, a team within Google, develops the PerspectiveAPI machine learning comment scoring system, which is used by numerous social media companies and publishers. Our system is based on distillation and uses a convolutional neuralnetwork to score individual comments according to several attributes using supervised training data Copyright ©2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). labeled by crowd workers. Note that PerspectiveAPI actually hosts a number of different models that each score different attributes.

[REF4] - paperID: 336e531a59cafbe215b950fd749bca866b89cea0 Title: SNK @ DANKMEMES: Leveraging Pretrained Embeddings for Multimodal Meme Detection (short paper) Chunk of text: For our task, we propose a fully automatic strategy based on a series of classifiers to assess style strength and content preservation. For style, we train a single classifier (main). For content, we train two classifiers that perform two ‘sanity checks’: one ensures that the two headlines (original and transformed) are still compatible (HH classifier); the other ensures that the headline is still compatible with the original article (AH classifier). See also Figure 1b. In what follows we describe these classifiers in237 (a) Overall data splits EVALUATION train & test main R+A3+A1 HH A1 + random pairs AH R+A3+A1 TASK train R+A3 test A2 (b) Training/test sets Figure 1: Data splits and their use in the different training sets more detail. When discussing baseline results, we will show how the contribution of each classifier is crucial towards a comprehensive evaluation. Main classifier The main classifier uses a pretrained BERT (Devlin et al., 2019) encoder with a linear classifier on top fine-tuned with a batch size of 256 and sequences truncated at 32 tokens for 6 epochs with learning rate 1e-05.

[REF5] - paperID: 336e531a59cafbe215b950fd749bca866b89cea0 Title: SNK @ DANKMEMES: Leveraging Pretrained Embeddings for Multimodal Meme Detection (short paper) Chunk of text: Finally, subtask 3 is the most challenging one, also because for some years only few training examples were available. More details on the document distribution in the training set are reported in Section 3. The aforementioned subtasks can be addressed in several ways. For example, researchers interested in historical content analysis can infer temporal information by looking at persons, places and time expressions, possibly integrating linking techniques. For those interested in studying semantic shifts, a purely lexical analysis may highlight changes in the lexical choices made by De Gasperi over time and give hints for document dating (Kulkarni et al., 2018). Also deep learning techniques, which proved effective on larger English corpora for document dating, could be tested (Vashishth et al., 2018). As an alternative, the subtasks could be addressed using document similarity techniques, so to assess to which training documents those in the test set are most similar, as-393 suming that similar documents have been written in the same years.

[REF6] - paperID: 336e531a59cafbe215b950fd749bca866b89cea0 Title: SNK @ DANKMEMES: Leveraging Pretrained Embeddings for Multimodal Meme Detection (short paper) Chunk of text: Elisa Bassignana, Valerio Basile, and Viviana Patti. 2018. Hurtlex: A Multilingual Lexicon of Words to Hurt. In Elena Cabrio, Alessandro Mazzei, and Fabio Tamburini, editors, Proceedings of the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018), Torino, Italy, December 10-12, 2018, volume 2253 of CEUR Workshop Proceedings. CEUR-WS.org. Christina Bosco, Felice Dell’Orletta, Fabio Poletto, Manuela Sanguinetti, and Maurizio Tesconi. 2018.

[REF7] - paperID: 336e531a59cafbe215b950fd749bca866b89cea0 Title: SNK @ DANKMEMES: Leveraging Pretrained Embeddings for Multimodal Meme Detection (short paper) Chunk of text: a bag-of-entities one. While this solution would be acceptable, and seemingly over the baseline according to the estimates on the validation set, it is reasonable to assume that the representations for these subtasks could be shared, improving the performances. Different variations of the same architecture are therefore evaluated on the validation set to monitor such improvement. In the final model, the sentence embeddings produced by SBERT are fed to a transformer block containing a multi-headed self-attention layer, its output is then averaged and concatenated with the bag-of-entities representation of the document before being fed to a multi-layer neural network. The output of each layer of this network is also fed to a dedicated neural network that produces the output of each subtask. The selected order for the subtasks in the multi-layer dense classifier places the historical classification first, followed by the five-years and then the single-year classification. A graphical representation of the architecture is in figure 2.

[REF8] - paperID: 7b8fe8c28a371120b4479540b2c8a0f7c5af25bf Title: Learned Text Representation for Amharic Information Retrieval and Natural Language Processing Chunk of text: The fourth orders ሃ /ha/, ሓ /ha/, ኃ /ha/, and ኻ /ha/ are normalized to ሀ /hə/, whereas the fourth orders ኣ /ʔa/ and ዓ /ʔa/ are normalized to አ /ʔə/. This is followed by sentence segmentation using “።” and tokenization into fundamental units (i.e., words, tokens, or n-grams) using space. Since pre-trained Amharic BERT tokenizers are unavailable publicly, we built WordPiece tokenizer in order to encode inputs for the MLM, NSP, and fine-tuned models. Previous studies in Amharic IR have shown the importance of word-based, stem-based, and root-based corpora for raw text representation . Here, we use the Amharic ad hoc information retrieval test collection (2AIRTC) corpus that also comes with the corresponding stem and root forms for learned text representation. The word-based, stem-based, and root-based documents were tokenized using the corresponding word-based, stem-based, and root-based WordPiece vocabularies. The tokenizers we used may segment words into multiple sub-words in order to more efficiently deal with out-of-vocabulary words and for better representation of complex words.

[REF9] - paperID: 336e531a59cafbe215b950fd749bca866b89cea0 Title: SNK @ DANKMEMES: Leveraging Pretrained Embeddings for Multimodal Meme Detection (short paper) Chunk of text: 5 Conclusion In the paper we described the approach proposed by the UNIGE SE team for the EVALITA 2020 PRELEARN shared task. The classifier relied on a set of features that was customised to address the specific requests of each sub-task. The results obtained by our models are all above baseline (if considered averaging the accuracies across all domains), although in some cases the results obtained by the baseline are still highly competitive. This suggests that automatic prerequisite learning is a difficult task requiring many different information to train the models. However, the obtained results suggest that, at least in a in-domain setting, 3https://github.com/mnarizzano/ se20-project-16 features extracted from raw texts are sufficient to achieve competitive results. In the cross-domain setting exploiting only this type of features is not enough. Nevertheless, using information extracted from knowledge structures allows to achieve better results in all sub-tasks.

........................................................................................................................................................................................................

Title: Exploring Multimodal Information Retrieval

Multimodal information retrieval (MIR) is a field of research that aims to improve the effectiveness of information retrieval systems by combining different modalities, such as text, image, audio, and video. In this section, we will explore some of the recent developments in MIR, focusing on the use of personalization, PageRank, and clustering techniques.

Personalization is a key component of MIR, as it allows the system to adapt to the user's preferences and needs. One approach to personalization is to use topic-sensitive PageRank, which takes into account the topic distribution over a document and the topic distribution over the query [REF0]. Another approach is to use ObjectRank, which is an authority-based keyword search model that assigns scores to objects based on their relevance to the query [REF1].

Clustering techniques are also widely used in MIR, as they can help to identify patterns and relationships between different modalities. For example, tensor spectral clustering has been used to partition higher-order network structures in MIR [REF5]. In addition, statistical approaches to mechanized encoding and searching of literary information have been proposed [REF2].

PageRank is another important component of MIR, as it can be used to rank the relevance of different documents based on their content similarity. Topic-sensitive PageRank has been combined with content similarity metrics such as the cosine similarity and the Jaccard index, resulting in improved retrieval effectiveness [REF3]. SeCoDeGrMa is another model that uses PageRank to improve retrieval effectiveness by incorporating Sub and Mul SMOs [REF4].

In conclusion, MIR is a rapidly evolving field of research that offers many opportunities for improving the effectiveness of information retrieval systems. Personalization, clustering techniques, and PageRank are just a few of the many approaches that are being explored in this field. As the amount of multimodal data continues to grow, it is likely that MIR will become an increasingly important area of research in the years to come.

References given to GPT:

[REF0] - paperID: d121c33a5a0d8b6615d8581cfee8a941ebc7daed Title: Graph-based entity-oriented search Chunk of text: Data source: • World Wide Web → Hyperlinks • Nearest Neighbors → Content Similarity Context: • Application → Search Components: • Personalization → Topic; Query • Personalization is not uniform per topic. • It is based on the topic distribution over a document. • And weighted by topic distribution over the query. • Normalized based on number of documents per topic. PageRank with content similarity Data source: • World Wide Web → Hyperlinks • Nearest Neighbors → Content Similarity Context: • Application → Search Components: • Personalization → Topic; Query • Adaptation of topic-sensitive PageRank.

[REF1] - paperID: d121c33a5a0d8b6615d8581cfee8a941ebc7daed Title: Graph-based entity-oriented search Chunk of text: “ObjectRank: AuthorityBased Keyword Search in Databases”. In: (e)Proceedings of the Thirtieth International Conference on Very Large Data Bases, Toronto, Canada, August 31 - September 3 2004. 2004, pp. 564–575. url: <http://www.vldb.org/conf/2004/> RS15P2.PDF (cit. on pp. 10, 57, 59, 71, 299, 335). A. Hogan, A. Harth, and S. Decker.

[REF2] - paperID: d121c33a5a0d8b6615d8581cfee8a941ebc7daed Title: Graph-based entity-oriented search Chunk of text: “A Statistical Approach to Mechanized Encoding and Searching of Literary Information”. In: IBM Journal of Research and Development 1.4 (1957), pp. 309–317. doi: 10.1147/rd.14.0309 (cit. on pp. 3, 5, 14, 31, 75, 226, 227). K. S. Jones. “A statistical interpretation of term specificity and its application in retrieval”. In: J. Documentation 60.5 (2004 ), pp. 493–502.

[REF3] - paperID: d121c33a5a0d8b6615d8581cfee8a941ebc7daed Title: Graph-based entity-oriented search Chunk of text: In this equation, Tjd represents the topic similarity between topic j and document d, PR0 jd represents the “balanced” PageRank for topic j and document d, and Qjq represents the similarity between topic j and query q based on the word distribution of all documents in topic j. The major difference between topic-driven and topic-sensitive PageRank is that now the probability vector for a topic is not uniform, and both topics and queries must surpass a given threshold of similarity with the topic, otherwise they won’t contribute to the score — this excludes less relevant documents and reduces topic drift. Rezvani and Hashemi combined the topicsensitive PageRank with content similarity metrics like the cosine similarity and the Jaccard index, experimenting with different query classifiers — naive Bayes and maximum entropy. Equation A.9 illustrates the main idea behind their proposal, showing how they calculate PageRank for a given topic T, represented by its documents, considering the similarity SimT (·, ·) between pages from topic T. Consider N−(·) and N+(·) as the set of incoming and outgoing pages, respectively. SimPRT (i) = 1 − d |T| + d X j∈N−(i) SimPRT (j) SimT (j, i) P k∈N+(j) SimT (j,k) (A.9) Notice that the PageRank of incoming pages is now distributed according to the normalized similarity over all their outgoing links instead of using the outdegree as is customary. They obtained the best results for the topic-sensitive PageRank when using the cosine similarity metric and naive Bayes to classify queries.

[REF4] - paperID: 06227bc74bcee55471fb37bde0149b317f8a2014 Title: Enhancing Semantic Code Search With Deep Graph Matching Chunk of text: The procedure is clearly depicted in Equation (6). • SeCoDeGrMa(Mul): This model is an alternative to Equation (7) that uses the Mul SMO. • SeCoDeGrMa: This model type concatenates two SMOs (i.e., Sub and Mul). Equation (8) demonstrates how this model works. Table 5 demonstrates that models incorporating SMOs (such as Sub, Mul, and SubMul) give significantly higher results 14 VOLUME 4, 2016 This article has been accepted for publication in IEEE Access.

[REF5] - paperID: d121c33a5a0d8b6615d8581cfee8a941ebc7daed Title: Graph-based entity-oriented search Chunk of text: 321–345. doi: 10.1137/16M1074023 (cit. on p. 310). A. R. Benson, D. F. Gleich, and J. Leskovec. “Tensor Spectral Clustering for Partitioning Higher-order Network Structures”. In: Proceedings of the 2015 SIAM International Conference on Data Mining, Vancouver, BC, Canada, April 30 - May 2, 2015. 2015, pp.

[REF6] - paperID: d121c33a5a0d8b6615d8581cfee8a941ebc7daed Title: Graph-based entity-oriented search Chunk of text: -0.4649 0.0147 0.1650 2010038 -0.7539 0.8086 0.0316 0.7475 2010040 -0.7740 0.9788 0.0000 0.8806 2010049 -0.7455 0.9460 0.0000 0.8324 2010057 -0.6500 0.9242 0.0526 0.8011 2010079 -0.7295 0.9730 0.0000 0.9382 2010096 -0.6864 0.8871 0.0000 0.7487 µ -0.7362 0.6889 0.0099 0.7228 σ 0.0541 0.5272 0.0183 0.2876 Table B.2 shows the average ρ values, hρ1i and hρ2i, for each experiment, per topic. It also shows the respective average Jaccard indexes, hJ1i and hJ2i, as a complement to correlation analysis. At the end of the table, mean (µ) and standard deviation (σ) values are shown to summarize global behavior. As we can see, when comparing TF-IDF and RWS, we obtained values for ρ1 that consistently approximate −1, with a mean of −0.7362 ± 0.0541, an indication that TF-IDF and RWS are anticorrelated. If we look at hJ1i, we find extremely low similarity values between the document sets returned by TF-IDF and RWS, with this value ranging around 0.0099 ± 0.0183. This explains the negative correlation and, interestingly, shows that RWS can still achieve good retrieval effectiveness while returning an almost completely different set of documents than TF-IDF.

[REF7] - paperID: d121c33a5a0d8b6615d8581cfee8a941ebc7daed Title: Graph-based entity-oriented search Chunk of text: Accessed on 2012-01-25. 2003 (cit. on pp. 72, 122). J. P. Callan, W. B. Croft, and S. M. Harding. “The INQUERY Retrieval System”. In: Proceedings of the International Conference on Database and Expert Systems Applications. Valencia, Spain, 1992, pp. 78–83 (cit.

[REF8] - paperID: d121c33a5a0d8b6615d8581cfee8a941ebc7daed Title: Graph-based entity-oriented search Chunk of text: url: <http://interjournal.org/manuscript_abstract.php?361100992> (cit. on p. 185). M. Bastian, S. Heymann, and M. Jacomy. “Gephi: An Open Source Software for Exploring and Manipulating Networks”. In: Proceedings of the Third International Conference on Weblogs and Social Media, ICWSM 2009, San Jose, California, USA, May 17-20, 2009. 2009. url: <http://aaai.org/ocs/index.php/> ICWSM/09/paper/view/154 (cit. on p. 185).

[REF9] - paperID: d121c33a5a0d8b6615d8581cfee8a941ebc7daed Title: Graph-based entity-oriented search Chunk of text: That is, the generation of related\_to hyperedges based on grouping by subject will only be complete after all documents in the collection are fully iterated over, while co-occurrence based hyperedges can be computed for each document, independently. Figure 7.3 illustrates the two approaches. In Figure 7.3a, we find two related\_to hyperedges, forming the sets {e1, e2, e3} and {e3, e4, e5, e6, e7}. As we can see, there is a certain redundancy in defining relations based on entity co-occurrence at the document-level, since the document hyperedge already ties the entities. Using sentence-level co-occurrence could be more interesting, however scaling would become an issue, given the much higher number of resulting hyperedges, not to mention the additional preprocessing overhead. It is however clear that the chance to visit nodes like e3 is still reinforced, as is the overall preference for entities as traversal nodes. In analogy to a Wikipedia research task about a given a subject, this would model the higher likelihood of a user following a link to another Wikipedia page (an entity), rather than issuing a new query based on terms extracted from the article.

........................................................................................................................................................................................................

Title: Exploring Different Textual Representations for Ranking in Neural Information Retrieval

In recent years, there has been a growing interest in exploring different textual representations for ranking in neural information retrieval. One of the most widely used pre-trained Language Models (PLMs) is BERT (Devlin et al., 2018) [REF0]. BERT is trained using the masked language modeling task that randomly masks some tokens in a text sequence, and then independently recovers the masked tokens by conditioning on the encoding vectors obtained by a bidirectional Transformer. In this work, we used a specific output from one of those layers to accomplish the calculation of a deep representation of the text. This operation can be expressed by: h0 = Hl0(texttok) hi = Hli(hi−1) [REF0].

Different ways of combining scores of a token from different sections have been tested to derive a full-text score for the token. In (Jimmy Lin, 2009) [REF1], the author found that computing the article score as the maximum score over all spans is superior to computing the score for an article as the sum of scores over all spans. In contrast, (Hearst & Plaunt, 1993) [REF1] found that using the sum of scores over all spans in scoring a document produces a superior ranking when evaluated on a dataset of 43 queries and 274 full-text documents.

The Aspect-Based Sentiment Analysis task is an extension of both the ATE and the SA tasks. The aim of the Aspect-Based Sentiment Analysis task is to detect the sentiment polarity associated with each aspect extracted, thanks to the ATE task. The sentiment score of every aspect detected is assumed to be the one associated with the portion of text in which it is contained. In order to do so, portions of the review are split using strong punctuation marks and some conjunctions. (Di Rosa and Durante, 2018) [REF2].

In the pre-processing stage, a root extraction algorithm was applied, and then a combination of Markov and fuzzy C-means was used for clustering. In the third stage, the DBN was used to build the Arabic classification model for each resulting cluster. The authors reported an F-score of 91.02%. Reference studied the effect of stemming techniques on Arabic document classification. For the pre-processing step, the authors picked three stemmers, one root-based, and two stem-based. (REF3).

The impact of the number of feedback documents and the 𝛽 parameter, which controls the emphasis of the expansion embeddings during the final document scoring, has been analyzed. When 𝑓𝑏 = 3, both Ranker and ReRanker obtain their peak MAP values. In addition, for a given 𝑓𝑏 value, the Ranker exhibits a higher performance than the ReRanker. Considering too many feedback documents causes a query drift, in this case by identifying unrelated embeddings. (REF4).

In dense retrieval, the documents and queries are represented using embeddings. Two distinct families of approaches have emerged: single representation dense retrieval and multiple representation dense retrieval. In single-representation dense retrieval, each query or document is represented entirely by the single embedding of the [CLS] (classification) token computed by BERT. Query-document relevance is estimated in terms of the similarity of the corresponding [CLS] embeddings. (REF7).

To alleviate the dependence of Neu-IR models on large-scale relevance supervision, leveraging weak supervision signals that are noisy but available at mass quantity has been proposed. Various weak supervision sources have been used to approximate query-document relevance signals, e.g., pseudo relevance labels generated by unsupervised retrieval methods, and title-document pairs. (REF9).

In conclusion, exploring different textual representations for ranking in neural information retrieval is an active area of research. Different approaches have been proposed, including the use of pre-trained Language Models, combining scores of a token from different sections, and leveraging weak supervision signals. These approaches have shown promising results and are expected to continue to advance the field of neural information retrieval.

References given to GPT:

[REF0] - paperID: bd23ce64a6422c1f73acf51675e53b7a06547da3 Title: UOBIT @ TAG-it: Exploring a Multi-faceted Representation for Profiling Age, Topic and Gender in Italian Texts Chunk of text: 3.2 Italian BERT Finally, we use a pre-trained BERT4 to accomplish the calculation of a deep representation of the text. One of the most widely used autoencoding pre-trained Language Models (PLMs) is BERT (Devlin et al., 2018). BERT is trained using the masked language modeling task that randomly masks some tokens in a text sequence, and then independently recovers the masked tokens by conditioning on the encoding vectors obtained by a bidirectional Transformer. Inside BERT, the information is passed forward crosswise transformer layers. In this work, we used a specific output from one of those layers, this operation can be expressed by: h0 = Hl0(texttok) hi = Hli(hi−1)

[REF1] - paperID: 27c12b8d9cfe4e88e513a53e620094e3a87a6ab2 Title: Measuring the relative importance of full text sections for information retrieval from scientific literature. Chunk of text: = !(#$%). Such scores for tokens can be added if the naive assumption of independence of the BM25 scores on which they are based is reasonably accurate. Now we test different ways of combining scores of a token from different sections to derive a fulltext score for the token. In (Jimmy Lin, 2009), the author found that computing the article score as the maximum score over all spans is superior to computing the score for an article as sum of scores over all spans. Spans in that work were paragraphs of full text documents from the TREC genomics collection, which consists of 36 topics (query questions) and manually annotated spans representing 2,477 full-text articles. In contrast, (Hearst & Plaunt, 1993) found that using the sum of scores over all spans in scoring a document produces a superior ranking when evaluated on a data set of 43 queries and 274 full text documents. Spans in (Hearst & Plaunt, 1993) are computed segments correlating with subtopics of a full text paper and are different from paragraphs.

[REF2] - paperID: bd23ce64a6422c1f73acf51675e53b7a06547da3 Title: UOBIT @ TAG-it: Exploring a Multi-faceted Representation for Profiling Age, Topic and Gender in Italian Texts Chunk of text: 4 Aspect-Based Sentiment Analysis The Aspect-Based Sentiment Analysis task is an extension of both the ATE and the SA tasks. In fact, the aim of the Aspect-Based Sentiment Analysis task is to detect the sentiment polarity associated to each aspect extracted, thanks to the ATE task discussed in Section 2. The possible polarity values are: Polarity Value neutral [0,0] positive [1,0] negative [0,1] mixed [1,1] Similarly to what we have done with the Aspect Category Polarity task at ABSITA 2018 (Di Rosa and Durante, 2018), we assumed that the sentiment score of every aspect detected in Section 2 is the one associated to the portion of text in which it is contained. In order to do so, we split portions of the review using strong punctuation marks and some conjunctions (especially the ones leading to sentiment inversion). For example, in the case of: Ottimo prodotto di marca, la qualita` e´ veramente notevole. Non e molto capi- ` ente ma si puo prendere un’altra ver-

[REF3] - paperID: cd52d4251de98217f32c3e556ea738ae97fc308d Title: Impact of Stemming and Word Embedding on Deep Learning-Based Arabic Text Categorization Chunk of text: A root extraction algorithm was applied in the pre-processing stage. Then, a combination of Markov and fuzzy C-means was used for clustering. In the third stage, the DBN was used to build the Arabic classification model for each resulting cluster. The experiment was conducted on 12,000 randomly selected documents from two different datasets. The authors reported an F-score of 91.02%. Reference studied the effect of stemming techniques on Arabic document classification. For the pre-processing step, the authors picked three stemmers, one root-based , and two stem-based , .

[REF4] - paperID: 44772b24ae2f68b77476c814b0607370f7195ddb Title: Pseudo-Relevance Feedback for Multiple Representation Dense Retrieval Chunk of text: However, if 𝐾 is too large, the returned embeddings contain more noise, and hence are not suitable for expansion – for instance, using 𝐾 = 64, feedback embeddings representing ‘innocent’ and ‘stunt’ are identified in Figure 2(b), which could cause a topic drift. Next, we analyse the impact of the number of feedback documents, 𝑓𝑏 . Figure 3(c) reports the MAP performance in response to different number of 𝑓𝑏 for both ColBERT-PRF Ranker and ReRanker. We observe that, when 𝑓𝑏 = 3, both Ranker and ReRanker obtain their peak MAP values. In addition, for a given 𝑓𝑏 value, the Ranker exhibits a higher performance than the ReRanker. Similar to existing PRF models, we also find that considering too many feedback documents causes a query drift, in this case by identifying unrelated embeddings. Finally, we analyse the impact of the 𝛽 parameter, which controls the emphasis of the expansion embeddings during the final document scoring.

[REF5] - paperID: 63483c9387d17e44eeb70c7321ad0dbb59b994fc Title: Universal Multimodal Representation for Language Understanding Chunk of text: Inspired by the previous success of visual-semantic embeddings, we apply neural image retrieval from the joint space to fetch a group of associated images. 3 UNIVERSAL REPRESENTATION FRAMEWORK This section overviews our universal representation framework. Given a sentence, we first fetch a group of matched images from our retrieval methods (details of our retrieval methods will be given in the next section). The text and images are encoded, respectively, by the text feature extractor and image feature extractor. Then the two sequences of representations are integrated using multi-head attention to form a joint representation, which is passed to downstream task-specific layers. Figure 1 overviews the whole multimodal representation model. 3.1 Encoding Layer 3.1.1 Text Encoder We pair each sentence with the top matched m images according to the retrieval method above.

[REF6] - paperID: 27c12b8d9cfe4e88e513a53e620094e3a87a6ab2 Title: Measuring the relative importance of full text sections for information retrieval from scientific literature. Chunk of text: Differences between all pairs of methods are statistically significant, except for the Max LogOdds and the Abs BM25 for the Set\_FT subset of PubMed Click Dataset. Based on these results we believe that log odds scoring is a useful approach for retrieval incorporating body text. The intuition behind it is that BM25 scores have a different meaning depending on the sections from which they are derived as illustrated in Fig 1. For a single query token, results in Figure 2 also suggest that the Sum scoring approach provides a better estimate of token importance than the Max scoring approach when using the log odds scoring for the Click dataset. If sections within a full text document were truly independent from each other, Sum LogOdds would be the ideal method to score a single query token over the multiple sections in a document. Figure 2. Average Precision for all query tokens is computed, averaged for each query and then over all queries for the PubMed Click dataset and its subset Set\_FT.

[REF7] - paperID: 44772b24ae2f68b77476c814b0607370f7195ddb Title: Pseudo-Relevance Feedback for Multiple Representation Dense Retrieval Chunk of text: Indeed, the BERT models have demonstrated further promise in being a suitable basis for dense retrieval. In particular, instead of using a classical inverted index, in dense retrieval, the documents and queries are represented using embeddings. Then, the documents can be retrieved using an approximate nearest neighbour algorithm – as exemplified by the FAISS toolkit . Two distinct families of approaches have emerged: single representation dense retrieval and multiple representation dense retrieval. In single-representation dense retrieval, as used by DPR and ANCE , each query or document is represented entirely by the single embedding of the [CLS] (classification) token computed by BERT. Query-document relevance is estimated in terms of the similarity of the corresponding [CLS] embeddings.

[REF8] - paperID: bd23ce64a6422c1f73acf51675e53b7a06547da3 Title: UOBIT @ TAG-it: Exploring a Multi-faceted Representation for Profiling Age, Topic and Gender in Italian Texts Chunk of text:   
 Figure 1: Scatter plot that shows the annotated score against the predicted score on the development set. We will now examine two reviews for which our regressor has the highest error. This is the text of the first review: “si autospenge proprio quando si necessita di usarla contelecomando”3 . This review was annotated with a score of 2, but the score assigned by our system is 4.75. This highlights a tendency of the system to give higher scores in uncertain cases. In this specific case we have no adjectives and two typing mistakes that result in no information from the lexicon and most of the words being disregarded as rare by our preprocessing pipeline. This suggests that a special treatment is needed for these specific cases where the classifier has fewer elements to take a decision.

[REF9] - paperID: 65c2d2ffe45569101860a7defc7cccbd36b3602a Title: Few-Shot Text Ranking with Meta Adapted Synthetic Weak Supervision Chunk of text: Moreover, they may often be overly confident and more unstable in the learning process (Qiao et al., 2019). A promising direction to alleviate the dependence of Neu-IR models on large-scale relevance supervision is to leverage weak supervision signals that are noisy but available at mass quantity (Zheng et al., 2019b; Dehghani et al., 2017; Yu et al., 2020). Through IR history, various weak supervision sources have been used to approximate querydocument relevance signals, e.g., pseudo relevance labels generated by unsupervised retrieval methods (Dehghani et al., 2017; Zheng et al., 2019b), and title-document pairs (MacAvaney et al., 2019). Recently, Zhang et al. (2020b) treat paired anchor texts and linked pages as weak relevance signals and propose a reinforcement-based data selection method ReInfoSelect, which learns to filter noisy anchor signals with trial-and-error policy gradients. Despite their convincing results, anchor signals are only available in web domains. Directly applying them to non-web domains may suffer from suboptimal outcomes due to domain gaps. To obtain weak supervision that adapts arbitrary domains, Ma et al.

........................................................................................................................................................................................................