

# Contextual Font Recommendations Based on User Intent

Sanat Sharma<sup>1,\*</sup>, Jayant Kumar<sup>2</sup>, Jing Zheng<sup>2</sup> and Tracy Holloway King<sup>2</sup>

<sup>1</sup>Adobe, Inc., 11501 Domain Drive, Suite 110, Austin, TX 78758 USA

<sup>2</sup>Adobe, Inc., 345 Park Ave, San Jose, CA 95110 USA

## Abstract

Adobe Fonts has a rich library of over 20,000 unique fonts that Adobe users utilize for creating graphics, posters, composites etc. Due to the nature of the large library, knowing what font to select can be a daunting task that requires a lot of experience. For most users in Adobe products, especially casual users of Adobe Express, this often means choosing the default font instead of utilizing the rich and diverse fonts available. In this work, we create an intent-driven system to provide contextual font recommendations to users to aid in their creative journey. Our system takes in multilingual text input and recommends suitable fonts based on the user's intent. Based on user entitlements, the mix of free and paid fonts is adjusted. The feature is currently used by millions of Adobe Express (shown in Fig. 1) users with a CTR of >25%.

## Keywords

font recommendations, intent understanding, text processing, neural networks, triplet learning, recommendation systems

## 1. Introduction: Application

Fonts, including the difficulty of font search and recommendations, have been a well-studied area in industry and at Adobe in particular. Previous works have tagged fonts with a small array of tags which can then be used for search. For example, Adobe Fonts ascribes 26 tags to fonts on its site. Recent works have used the MyFonts dataset to assign tags to fonts and utilized them for retrieval [1] [2]. There are also projects that perform font pairing and retrieval based on visual appearance of the font [3] [4].

However, doing font retrieval based on font tags often misses the semantics governing what types of contexts the font should be used in. Furthermore, retrieving and recommending fonts based on the visual style of the fonts already on the canvas often leads to poor results when there is variety in the content of the page.

Our work focuses on understanding the intent of the the user's project using multimodal input (i.e. text and image context). We then recommend fonts that work well with the project

---

*eCom'23: ACM SIGIR Workshop on eCommerce, July 27, 2023, Taipei, Taiwan*


\*Corresponding author.

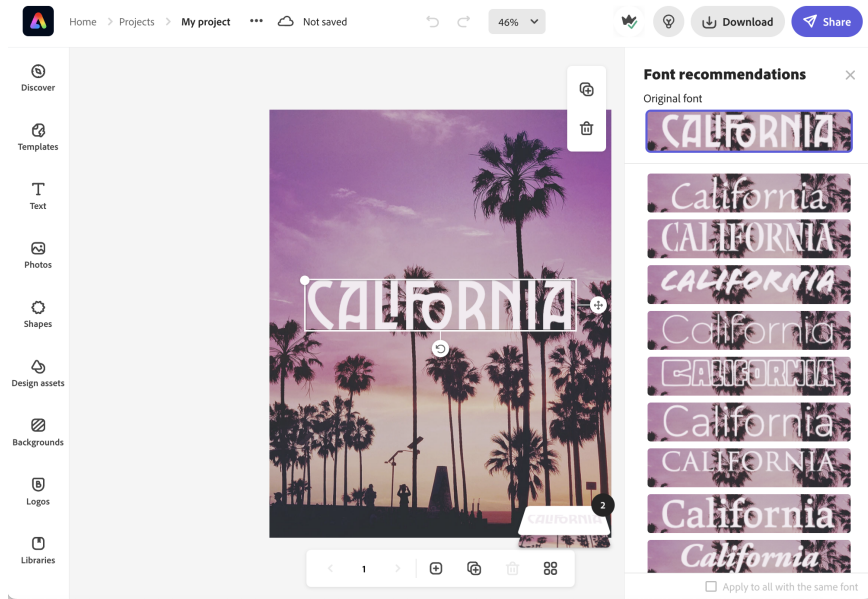
✉ sanatsha@adobe.com (S. Sharma); jaykumar@adobe.com (J. Kumar); jinzheng@adobe.com (J. Zheng); tking@adobe.com (T.H. King)

🆔 0009-0003-2479-3041 (S. Sharma); 0009-0004-3858-5961 (J. Kumar); 0009-0004-6097-4146 (J. Zheng); 0000-0002-7956-505X (T.H. King)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)



**Figure 1:** Example font recommendations in Adobe Express.

intent. This allows us to understand the semantic relationship between the user’s input (e.g. the text they need to select a font for) and the fonts in our library. Our work comprises:

1. A multimodal Input → Font model that can accurately and reliably recommend a diverse array of fonts to the user, by extracting their creative intent.
2. A method to recommend “similar” fonts by intent, rather than simple visual appearance.
3. A scalable query architecture for low latency recommendations.

Since we map the user’s context to intent and then to fonts, we are able to showcase a diverse set of fonts, including specialized, unique fonts. This intent-based method also allows us to train our models on a much smaller set of data, which is often a problem in training font recommendation models. The approach is currently used in production in Adobe Express, with a CTR of >25% and a downstream project export rate of >50% whenever a recommended font has been clicked: that is, users who interact with the recommended fonts are likely to finish and export the project they are working on. Whether the user is using a free, trial, paid, or enterprise account has an effect on engagement with font selection in general and with subsequent use of the font recommendations.

## 2. Data

In order to mine data for training, we extracted text-font pairs for ~212,000 texts in Adobe Express templates. Adobe Express is a platform for people to create flyers, posters, cards, logos, etc. and hence has templates with a diverse set of fonts. Previous font recommendation

approaches often utilized general text documents, which rarely have much diversity of fonts [5] [3].

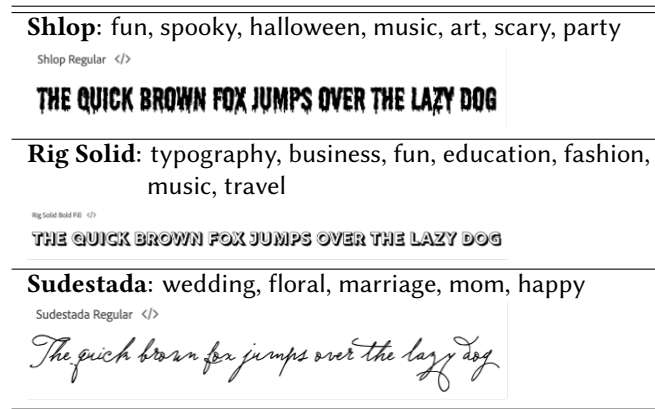
For each text input, we determine the top intents via our intent detection model. The intent detection model is trained on 1500+ intents, derived from Adobe Express template tags. These tags range from events such as “Christmas” and “Halloween” to objects and emotional characteristics such as “balloon”, “happy”, and “encouraging”. This results in triplets of font + text + intent. This, to our knowledge, is the largest dataset of font-text-intent mappings. Tab. 1 showcases a sample of the training dataset. After further thresholding and preprocessing, we extract 470,000 rows of text + font + intent triplets that we utilize to train our models.

**Table 1**

Training dataset of texts, intents and font.

| Id  | Text        | Font          | Intent       | Score  |
|-----|-------------|---------------|--------------|--------|
| 150 | photography | Gizmo         | photography  | 0.0249 |
|     |             |               | camera       | 0.0063 |
|     |             |               | photographer | 0.0038 |
|     |             |               | woman        | 0.0020 |
|     |             |               | sunset       | 0.0018 |
| 693 | Manufactory | Myriad Hebrew | food         | 0.0021 |
|     |             |               | business     | 0.0013 |
|     |             |               | restaurant   | 0.0013 |
|     |             |               | cafe         | 0.0007 |

Since fonts within a font family are visually similar and have other similar characteristics, we group similarly used fonts in a font family together based on intent. This grouping helps capture artistic factors of the fonts and corresponding intents. The final training set consists of 2043 unique font families to recommend from. Fig. 2 shows examples of fonts and their associated intents.



**Figure 2:** Font characteristics based on intent: The font name is followed by the related intents (from most to least relevant). An example of the font is shown below the font name.

## 3. Model

### 3.1. Intent Detection Model

#### 3.1.1. Creative Intent Taxonomy

In order to map user creative intent, we semi-automatically extract a creative intent taxonomy from the Adobe Express content metadata and users' behavioral data when using and searching for content. We utilize Adobe Express templates for this taxonomy since the template data contains rich designer-added tags and represents a diverse set of creative intents. The intents comprise two main types (example broad and narrower intents in parentheses):

1. Topic of work/task/emotion (e.g. yoga, Halloween, cosmetic, food, book launch, happiness, birthday, birthday party)
2. Type of creation (e.g. flyer, poster, social media post, card, greeting card, christmas card, valentine's day card)

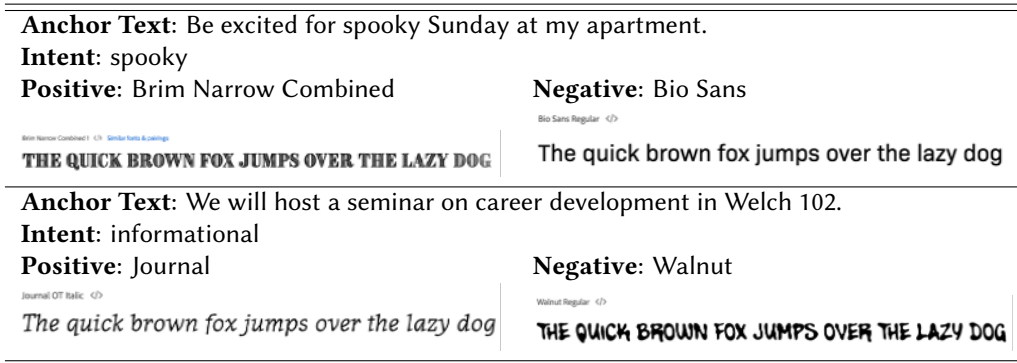
We utilize a pretrained DistilBert [6] transformer to cluster related topics and then apply popularity metrics to derive a creative intent taxonomy of over 1500 intents mined from the Adobe Express dataset. This clustering allows us to collapse terms such as "mlk jr day", "martin luther jr day" and "mlk day" into one concept. The most popular term (most frequent occurrence among all templates) is used to represent the intent.

#### 3.1.2. Multilingual Text-Intent Model

Users often add text to their creation to convey their intent. For example, "Happy Mother's Day" is written on Mother's Day greeting cards, and "We stand in solidarity with black businesses" for BLM social media posts. It is crucial to infer the intent (topic + type) of a user project based on the text entered on the canvas or issued as a search. We trained a transformer model which takes as input a text sample (sentence or phrase or word) and predicts the top, most probable intents from our curated taxonomy. We extracted over 335,000 texts from Adobe Express templates and mapped them to the "topics" that were manually tagged by the template designers. Our final dataset consists of over 1 million text-intent pairs that we utilized to train our classification model. We fine-tuned a DistilBert [6] using contrastive learning to learn the text-to-intent relationships. Our model works for 36+ languages including English, French, German and Japanese, which is crucial for any production feature at Adobe.

### 3.2. Font Recommendation Model

In order to model intent and fonts in the same embedding space efficiently, we utilized triplet learning and online triplet mining [7]. This effectively generates intent-to-font pairings, as shown in Fig 3. We fine-tuned a pretrained SimCSE model trained on query understanding [8] and trained the model with Triplet Margin Loss [9] with a margin of 2. The loss function is summarized in equation 1 where:  $\alpha$  is the margin;  $A$ ,  $P$  and  $N$  denote the anchor, positive and negative sample respectively;  $f$  denotes the neural network. We utilize Adam [10] as our



**Figure 3:** Triplet pairs used for training. We use intents and positive, negative fonts for training.

optimization function with a high learning rate (empirically we found  $3e^{-3}$  to give the best results).

$$L(A, P, N) = \max(\|f(A) - f(P)\|^2 - \|f(A) - f(N)\|^2 + \alpha, 0) \quad (1)$$

One challenge was to prevent commonly used fonts from skewing recommendation variation. For generating the embedding for each font, we found representing a font by a ranked set of its top 7 intents gave the best results. The ranking is based on the frequency of the intent-font pair in our dataset. This helped distinguish unique fonts and remove font bias at runtime.

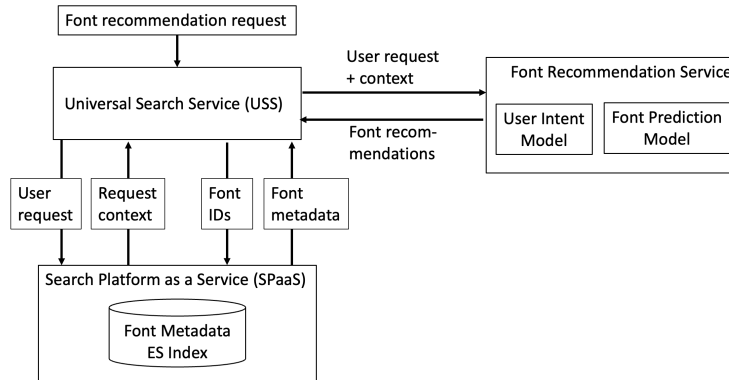
## 4. Service Architecture

The system design of font recommendation has three main components: the font recommendation service, Search Platform as a Service (SPaaS) and Universal Search Service (USS).

The USS service defines the input-output schema for clients and acts as the orchestrator between the font recommendation service and the search platform (SPaaS).

SPaaS contains the metadata, embeddings and intent information across various Adobe assets such as backgrounds, images, fonts and Adobe Express templates. SPaaS is responsible for providing the font recommendation service with input information (context, embeddings, etc.) present in its Elasticsearch (ES) index as well as extracting font metadata from the ES index for each recommended font returned from the service.

At runtime, our clients send user requests (i.e. text input from the Adobe Express editing canvas) to USS. USS retrieves additional information from SPaaS before sending the full payload to the font recommendation service. The font recommendation service extracts the intents from the text input and then uses the weighted set of intents to generate an embedding. This embedding is used to select font recommendations. When the input language has a restricted set of compatible fonts (e.g. Japanese), SPaaS provides the recommendation service with the subset of fonts to recommend from. The full architecture is shown in Figure 4.



**Figure 4:** Service Architecture for Font Recommendation

## 5. Evaluation

### 5.1. External Manual Evaluation

We conducted an external evaluation with 75 evaluators. We selected 78 high frequency templates and displayed them to annotators with one of the texts removed. We then showed the annotators 5 recommended fonts for the text that had been removed and asked them to rate each of the fonts as “very good”, “ok” or “not good”. Since font recommendations are highly subjective, we asked the annotators to fully consider each font’s compatibility with the text. We found that evaluators considered a text-font pair relevant (very good or ok) 81% of the time (Table 2).

We also evaluated a weak baseline of randomly selected fonts and a strong baseline of common fonts. Both were  $>10$ ppt (see fn. 2) worse than the intent-based algorithm. Randomly selected fonts often showed inappropriate, heavily stylized fonts. Popular fonts showed visually similar fonts with little diversity or newness: This reduced discoverability of unique fonts (e.g. for Halloween, where stylized fonts work well, only non-stylized options were shown).

### 5.2. Internal Manual Evaluation

We also conducted an internal evaluation with 16 annotators before AB testing the font recommendations. Similar to the external evaluation, we took 100 texts in multiple languages (English, French, German, Japanese) from Adobe templates and asked team members to rank the font recommendations on a scale of 1–5 (1 = very bad, 5 = very good). Our internal results reflected the external evaluation, with a mean score of 3.67. Since fonts are relatively subjective, this was considered sufficient for AB testing. The score distribution is shown in Table 2. An example internal evaluation query is shown in Figure 5.

**Table 2**

Relevancy results for internal & external evaluations.

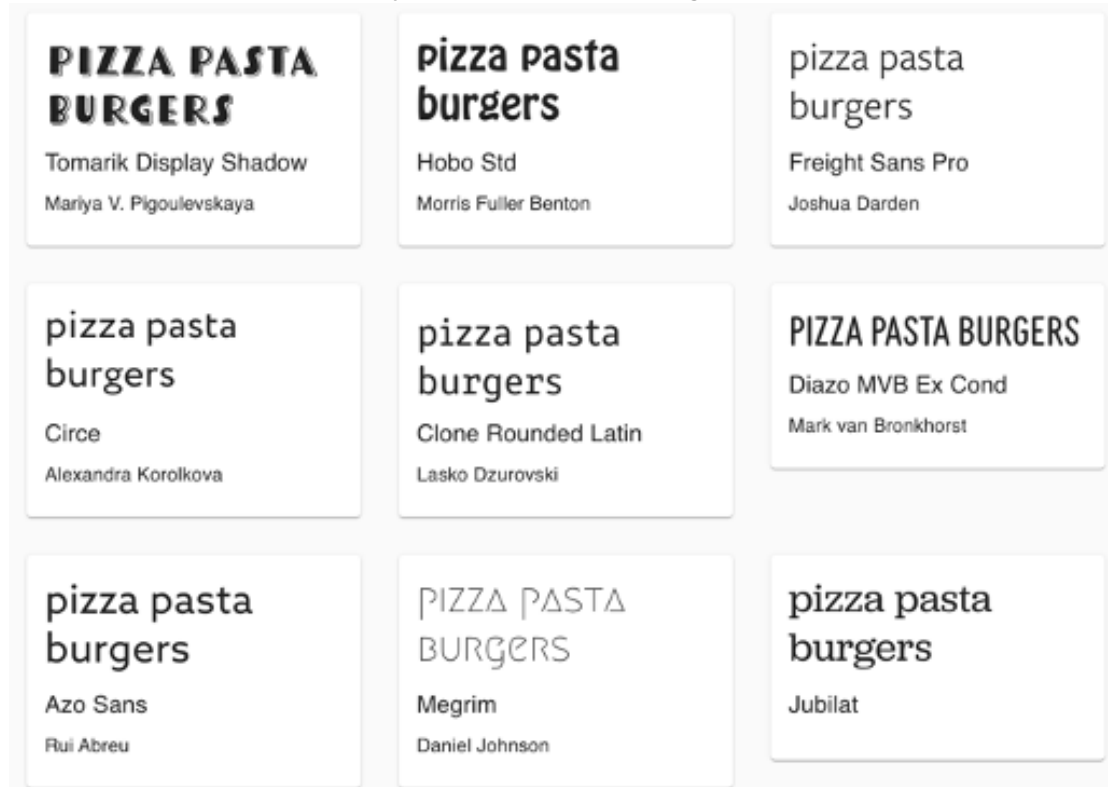
| External evaluation: 75 annotators |     |          |  |  |
|------------------------------------|-----|----------|--|--|
| Very Good                          | Ok  | Not Good |  |  |
| 464                                | 216 | 159      |  |  |

| Internal evaluation: 16 annotators |      |    |     |          |
|------------------------------------|------|----|-----|----------|
| Very Good                          | Good | Ok | Bad | Very Bad |
| 10                                 | 65   | 56 | 11  | 1        |

**Text:** pizza pasta burgers

**Extracted Intents:** Food, takeaway, fast food, restaurant, burger



**Figure 5:** Top Text-to-Font recommendations for the text **pizza pasta burgers**. Classic menu fonts for casual restaurants are recommended.

### 5.3. In-Production Analysis

Once the external and internal evaluations were completed, the font recommendations were put in production<sup>1</sup> in Adobe Express in the experience shown in Fig. 1. Overall, the recommendations have a click-through rate (CTR) of >25% and users who select a recommended font have a downstream project export rate of >50%.

<sup>1</sup>We cannot report exact numbers for proprietary reasons. Instead we provide lower bounds and relative numbers.

There are 3 types of user accounts on Adobe Express: Free, trial, and enterprise. Although the font recommendation algorithm is identical for all users, the percentage of free and paid fonts depends on their account type. We analyzed the engagement and success of the different types, focusing on users who edited text since font recommendations are only relevant to such users. Surprisingly, free and enterprise users who edited text had the same rates for using the font selection module, selecting a recommendation, and exporting a final project. In contrast, trial users had much higher engagement at each step (font module +15ppt; recommendation selection +13ppt; export +2ppt).<sup>2</sup>

Users who utilize font recommendations have a higher project export (+25ppt) compared to users who do not. This may be due to the fact that engaged users have higher desire to complete their projects. This is also consistent with overall project export rates after click on font recommendation for paid (trial + enterprise + non-enterprise paid) vs. free users. Paid users had a +10ppt export rate compared to free users after clicking on font recommendations.

We also saw much higher (3x) engagement on web compared to mobile surfaces. We attribute this to the fact that the recommendation feature is more prominent on web and that having a larger screen allows users to browse recommendations more easily.

Finally we draw some key insights about paid vs. free fonts. For paid users, where there is an equal split of free and paid fonts, users click on paid fonts slightly more (+2ppt) than free ones. In contrast, free users are shown a much higher percentage of free fonts and rarely select paid fonts due to the payment barrier. This suggests that while paid fonts are generally better liked, they are not enough incentive for most users to convert to paying customers.

## 6. Conclusion and Next Steps

This work showcased a novel end-to-end framework for recommending fonts based on multilingual user input. The recommendation system includes a model to understand user intent and brings intents and fonts into the same representation space. It includes a low-latency, scalable architecture to serve the font recommendations. The current approach has high accuracy and relevance in its recommendations and is in production in Adobe Express.

Next steps focus on whole document relevance. Firstly, we want to understand the fonts already in the document so that we can recommend fonts that work in harmony with the existing artistic design. Secondly, the model described here focuses on individual texts, whereas global intent is not taken into consideration. Finally, we currently do not support right-to-left languages like Arabic.

## Acknowledgments

We thank Andrei Stefan, Ravindra Sadaphule, and Arpita Agrawal for their invaluable help in productionalizing the font recommendation feature shown in Fig. 1.

---

<sup>2</sup>We are unable to publish the exact rates. We report the results as percentage point (ppt) change. For example, a +5 in CTR indicates going from a CTR of X% (e.g. 60%) to a CTR of X+5% (e.g. 65%).



## References

- [1] T. Chen, Z. Wang, N. Xu, H. Jin, J. Luo, Large-scale tag-based font retrieval with generative feature learning, in: IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, 2019. URL: <https://arxiv.org/abs/1909.02072>. doi:10.48550/ARXIV.1909.02072, available on arXiv.
- [2] H. S. Baird, G. Nagy, Self-correcting 100-font classifier, in: SPIE-The international Society for Optical Engineering, Document Recognition, 1994, pp. 106–115.
- [3] S. Jiang, Z. Wang, A. Hertzmann, H. Jin, Y. Fu, Visual font pairing, CoRR abs/1811.08015 (2018). URL: <http://arxiv.org/abs/1811.08015>. arXiv:1811.08015.
- [4] S. Matsuda, A. Kimura, S. Uchida, Impressions2Font: Generating fonts by specifying impressions, CoRR abs/2103.10036 (2021). URL: <https://arxiv.org/abs/2103.10036>. arXiv:2103.10036.
- [5] F. Kadner, Y. Keller, C. A. Rothkopf, AdaptiFont: Increasing individuals' reading speed with a generative font model and Bayesian optimization, CoRR abs/2104.10741 (2021). URL: <https://arxiv.org/abs/2104.10741>. arXiv:2104.10741.
- [6] V. Sanh, L. Debut, J. Chaumond, T. Wolf, DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter, 2019. URL: <https://arxiv.org/abs/1910.01108>. doi:10.48550/ARXIV.1910.01108.
- [7] E. Hoffer, N. Ailon, Deep metric learning using triplet network, in: Similarity-Based Pattern Recognition: Third International Workshop, SIMBAD 2015, Springer International Publishing, 2015, pp. 84–92. URL: <https://arxiv.org/abs/1412.6622>. doi:10.48550/ARXIV.1412.6622, available on arXiv in 2014.
- [8] T. Gao, X. Yao, D. Chen, SimCSE: Simple contrastive learning of sentence embeddings, 2021. URL: <https://aclanthology.org/2021.emnlp-main.552>. doi:10.18653/v1/2021.emnlp-main.552.
- [9] D. P. Vassileios Balntas, Edgar Riba, K. Mikolajczyk, Learning local feature descriptors with triplets and shallow convolutional neural networks, in: E. R. H. Richard C. Wilson, W. A. P. Smith (Eds.), Proceedings of the British Machine Vision Conference (BMVC), BMVA Press, 2016, pp. 119.1–119.11. URL: <https://dx.doi.org/10.5244/C.30.119>. doi:10.5244/C.30.119.
- [10] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014. URL: <https://arxiv.org/abs/1412.6980>. doi:10.48550/ARXIV.1412.6980.