Corpus-based Automatic Text Expansion

Balaji Vasan Srinivasan[1], Rishiraj Saha Roy[2], Harsh Jhamtani[3], Natwar Modani[1], Niyati Chhaya[1]

[1]*Adobe Research Big Data Experience Lab, Bangalore, India*

[2]*Max Planck Institute for Informatics, Saarland Informatics Campus, Germany*

[3]*Language Technology Institute, Carnegie Mellon University, Pittsburgh, PA USA*

Content Authoring & Velocity

Content authors or copywriters are creative individuals

Several variations of same content required with limited resources

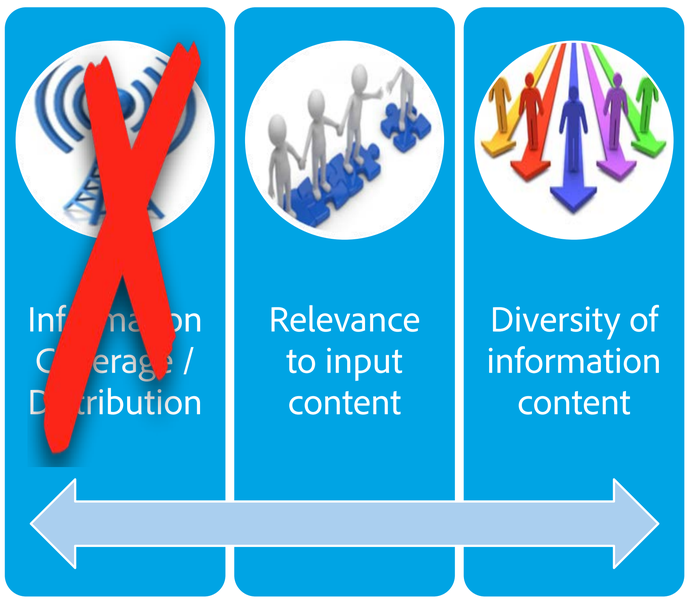
Diversification and personalization key to user engagement

The Problem of Text Expansion

Automatically expand a piece of textual content to a desired size

- By adding additional content from a repository
- Ensuring relevance of the overall content
- Providing information diversity in the constructed content
- *Accounting for overall coherence in the content*

Difference from Automatic Text Summarization [1]



Proposed Solution Framework

Input: content snippet → Retrieve relevant pieces of content from your repository → Compute the relevance of content to the snippet → Greedy algorithm to assemble content

Use of blur gallery in Photoshop

Blur Gallery + **Photoshop**

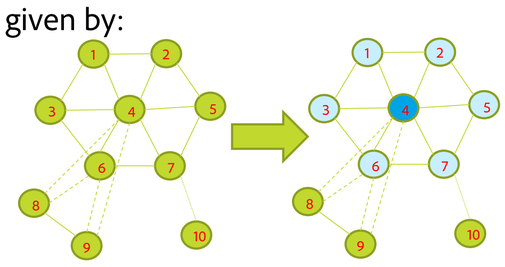
Blurs that set your images in motion. Want to create a sense of motion in your images? Easily add circular, elliptical, and path blurs with new Blur Gallery effects. See how it works **0.8**	Use Path Blur to add blur along any path and Spin Blur to create circular or elliptical blurs. The Mercury Graphics Engine makes all Blur Gallery interactions fast and fluid. See how it works **0.7**
Up-sample images up to 15 times faster (depending on file size and video card configuration) now that the Mercury Graphics Engine delivers an OpenCL performance boost. The engine powers new Blur Gallery motion effects and the Focus Mask feature, too. **0.5**	Apply Blur Gallery and Liquify effects nondestructively thanks to Smart Object support. Your original file stays intact as you add blur effects or push, pull, pucker, or bloat the image or video. Edit or remove the effects at any time even after saving your file. See how it works **0.6**
Stay up to date with instant access to new Photoshop features as soon as they're released. See whats new! **0.4**	

- Alternative1: Maximum Marginal Relevance [2]
  - Identify paragraph from the corpus R that satisfies:

$$arg\ \max_{D_i \in R \setminus S} [\lambda(Sim_1(D_i, Q)) - (1-\lambda)(\max_{D_j \in S} Sim_2(D_i, D_j))]$$

  - Update the selected set S and repeat until length/information criteria satisfied

- Alternative 2: Graph based Rewards [3]
  - Paragraph ::: Nodes in a graph, node reward = relevance of paragraph to input content
  - Edge Weight = similarity between paragraphs
  - Select paragraph with maximum Gain-Cost Ratio, Gain given by:
  - $G_{v_i}^l = r_i^{l-1} + \sum_{v_i \in N_i} w_{ij} r_j^{l-1}$
  - Update rewards of node as: $r_j^l = (1 - w_{i^* j}) r_j^{l-1}$
  - Repeat until length/information criteria satisfied

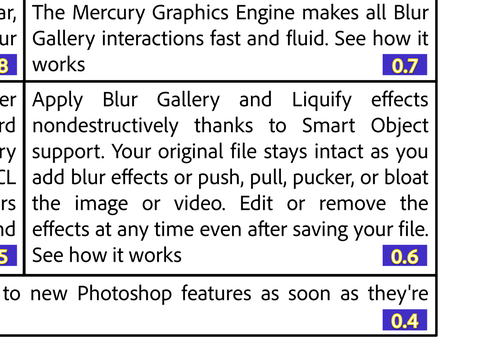


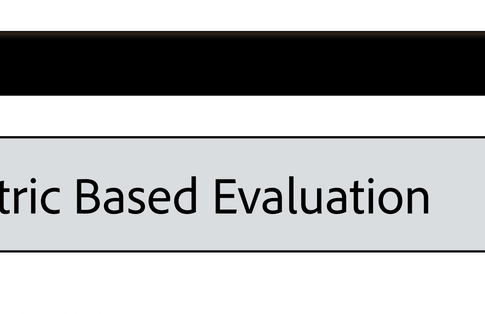
Human Annotation & Metric Based Evaluation

- Dataset: 215 proprietary forum articles around key product features and troubleshooting instructions
- Input: Constructed 30 short snippets (~35 words per snippet)
- Human Annotations:
  - 30 annotators – each evaluating 4 articles
  - Scoring Relevance & Diversity on a scale of 0 – 7
- Metric Based evaluation

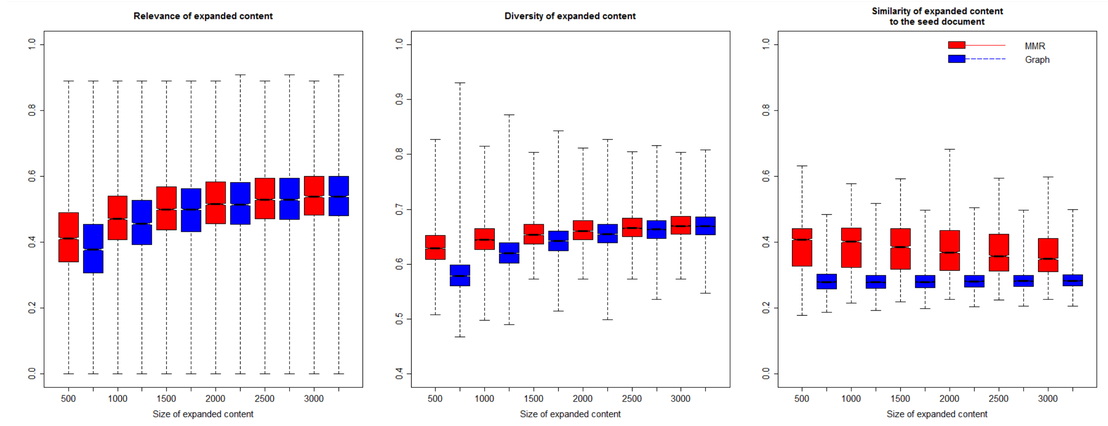
- Relevance: $\sum_{i=1}^{N} \left(\frac{1}{N} \sum_{j=1}^{N} \frac{\sum_{k=1}^{topK(ti,tj)} \lambda^k sim(ti,tj)}{\sum_{k=1}^{K} \lambda^k} \right)$

- Diversity: $\frac{1}{N} \sum_{i=1}^{N} \left(1 - \frac{1}{N} \sum_{j=1}^{N} \frac{\sum_{k=1}^{topK(ti,tj)} \lambda^k sim(XFi,XFj)}{\sum_{k=1}^{K} \lambda^k} \right)$





Evaluation on Australian Legal Dataset [4]



Seed	MMR Based Expansion	Graph Based Expansion
Damages claimed from respondents for breach of guarantee of profit shortfall. First respondent had ostensible authority to bind second respondent to oral variation. Profit shortfall amount for 1998 contracts evidence agency.	However it became clear during cross-examination of Mr Forbes and Mr Brauer that the sales which the respondents claimed should have been credited to the 1998 year actually took place in 1997, and were properly accounted as 1997 sales, as claimed by the applicants. In summary, the respondents claimed that these documents were critical to properly investigating: It was not in contention between the parties that the source financial documents were missing and unavailable. Did Forbes Australia experience a profit shortfall in the financial year ending 31 December 1998?	Did Forbes Australia experience a profit shortfall in the financial year ending 31 December 1998? The material is relevant to both the applicants' claims concerning the 1998 profit shortfall and the respondents' defense. 2. a claim for $1,691,284 which is alleged to be the profit shortfall in respect of the 1999 calendar year. It also follows that the thirty-eighth and thirty-ninth respondents should recover judgment for breach of duty. That shortfall was claimed in the amount of $71,663.65.

- Australian Legal Case Reports dataset: 3890 legal cases from the Federal Court of Australia
- Every case included a gold standard summary for every case in the form of `catchphrases' and `key sentences'
  - Used as the seed for expansion from the repository
- Future Exploration: Coherence of the expanded content

References

1. Nenkova, Ani, and Kathleen McKeown. "Automatic summarization." *Foundations and Trends® in Information Retrieval* 5.2–3 (2011): 103-233.
2. Carbonell, Jaime, and Jade Goldstein. "The use of MMR, diversity-based reranking for reordering documents and producing summaries." *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 1998.
3. Modani, Natwar, et al. "Creating diverse product review summaries: a graph approach." *International Conference on Web Information Systems Engineering*. Springer International Publishing, 2015.
4. http://archive.ics.uci.edu/ml/datasets/Legal+Case+Reports