Corpus-based Automatic Text Expansion

Balaji Vasan Srinivasan¹, Rishiraj Saha Roy², Harsh Jhamtani³, Natwar Modani¹, Niyati Chhaya¹

¹Adobe Research Big Data Experience Lab, Bangalore, India
²Max Planck Institute for Informatics, Saarland Informatics Campus, Germany
³Language Technology Institute, Carnegie Mellon University, Pittsburgh, PA USA

Content Authoring & Velocity
The Problem of Text Expansion
Difference from Automatic Text Summarization

Automatically expand a piece of textual content to a desired size

1. Alternative: Maximum Marginal Relevance [2]
   - Identify paragraph from the corpus R that satisfies
     \[
     \arg \max_{\theta \in D} \{ \theta \} = \arg \max_{\theta \in \Delta} \{ \theta \} - (1 - \lambda \theta \in \Delta) \]
   - Update the selected set S and repeat until length/information criteria satisfied

2. Alternative: Graph based Rewards [3]
   - Paragraph: Nodes in a graph, node reward = relevance of paragraph to input content
   - Edge Weight = similarity between paragraphs
   - Repeat until length/information criteria satisfied

Proposed Solution Framework

Use of Blur Gallery in Photoshop
Blur Gallery
Photoshop

Blur Gallery delivers an OpenCL performance boost. The engine powers Gallery effects. See how it works.

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.

Top-sample images up to 15 times faster (depending on file size and video card configuration) now that the Mercury Graphics Engine delivers on OpenCL performance boost. The engine powers new Blur Gallery motion, effects, and the Focus Mask feature,

Stay up to date with instant access to new Photoshop features as soon as they’re released. See what’s new!

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:
  - 3D annotations – each evaluating 4 articles
  - Scoring Relevance & Diversity on a scale of 1–7
- Metric Based evaluation

Evaluation on Australian Legal Dataset [4]

- Australian Legal Case Reports dataset with legal cases from the Federal Court of Australia
- Every case included a gold standard summary for every stance in the form of catchphrases and key sentences
- Used as the second experiment for the repository
- Future Directions: Consideration of the expanded content

References

© 2017 Adobe Systems Incorporated. All Rights Reserved.