

Technical Disclosure Commons

Defensive Publications Series

May 31, 2019

PDF CONTENT RE-USE EVALUATION AND FEEDBACK WHILE RIPPING FOR COMMON PRINTING

HP INC

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation

INC, HP, "PDF CONTENT RE-USE EVALUATION AND FEEDBACK WHILE RIPPING FOR COMMON PRINTING",
Technical Disclosure Commons, (May 31, 2019)
https://www.tdcommons.org/dpubs_series/2243



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.

PDF content re-use evaluation and feedback while ripping for commercial printing

Disclosed is a method for evaluating the performance implications of reusable content in a PDF file while ripping. Processing data is collected to allow the system to intelligently determine if it should switch PDF processing technics and present feedback to the user about the file being worked on.

Ripping the same content repeatedly in a PDF file can be a waste of resources and cause performance issues while ripping. Object reuse ripping can increase performance with this type of file by lowering the resources required and only ripping unique content and reused content only once. If no reusable objects are found in the PDF, then the system begins to waste time looking for reusable object and falls back to ripping each page uniquely. What a person sees as reusable content in a PDF file may not be what the system evaluating it considers reusable. Often ripping engines try to look for items that are not only reusable but reach a certain threshold of difficulty or size to make it worth the effort to cache and manage its content.

Active feedback while ripping gives a benefit of both worlds. You can start looking for reusable objects and determine if it is seeing any to determine if the ripping process should be switched in mid processing of the PDF file.

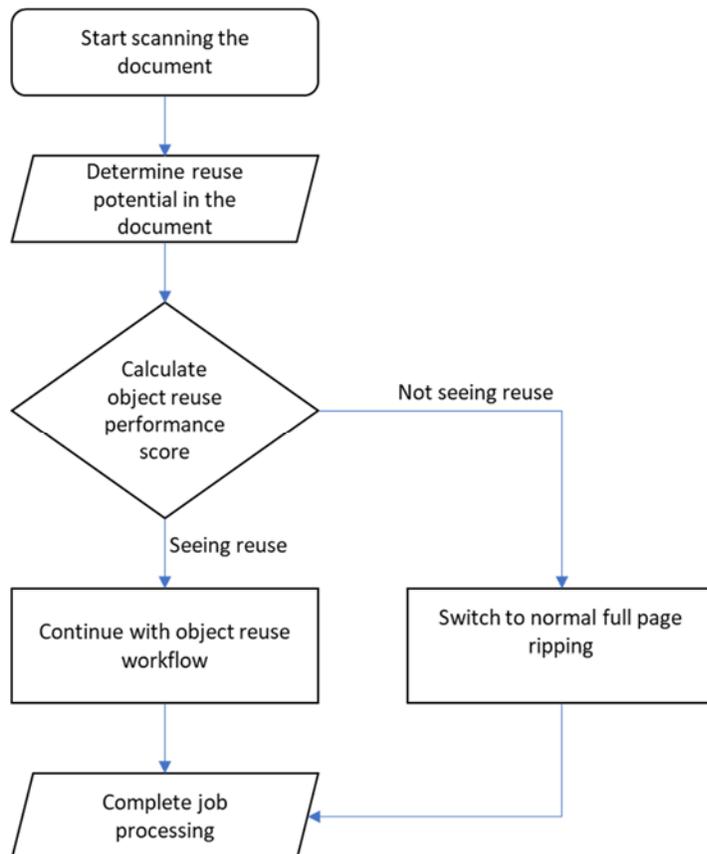


Fig. 1

This allows the optimization of performance for the given PDF file and the same data used to determine this can be presented to the user for feedback, giving them the option to not use this workflow for future jobs of similar file design or structure.

Performance tracking with object reuse ripping is evaluated into five metrics, we call affecting factors.

Percent Re-Used

This metric is looking at the percentage of the page that is covered with reusable objects vs the area covered by unique content. The higher the number the greater the performance boost we expect from the system. Difficulty of ripping the content is not part of this value.

Percent Page Coverage

This value corresponds to the total area of all the objects in a page compared to the actual dimension of the page. The overlapping of object ripped independently will affect overall performance. Percentage values over 100 will negatively affect the performance of the job.

Ripping

This is a calculated number based on time spend ripping content and Recomposition time for each page. When we detect hard to rip reusable content, then this metric show significant performance improvements. It reflects that we are saving significant time by reusing the ripped content. On the other hand, if we see rip time very low for easy to rip content, then the overhead of managing the object cache cuts into any performance improvements and may even slow the system down.

Object Count

This is the number of re-usable object per page. Large number of small items will increase page Recomposition overhead. By default, ripping engines typically coalesce smaller objects into bigger raster objects to improve the overall performance. This value will reflect if the system is able to coalesce the PDF reusable objects to improve performance.

Scan Time

This is the time spent by the rip engine to determine what is reusable and what is unique content on each PDF page. Scan time is a necessary process before the system can rip any object on the PDF page. The shorter the scan time the better for performance by allow the rip engine to get started more quickly.

These five metrics are presented to the user in a simple good, bad or neutral rating. Each of the five affecting factors has a scale factor associated with them that helps to give an overall performance score ranging from -5 to 5 to represent if object reuse ripping of the PDF was beneficial or not.

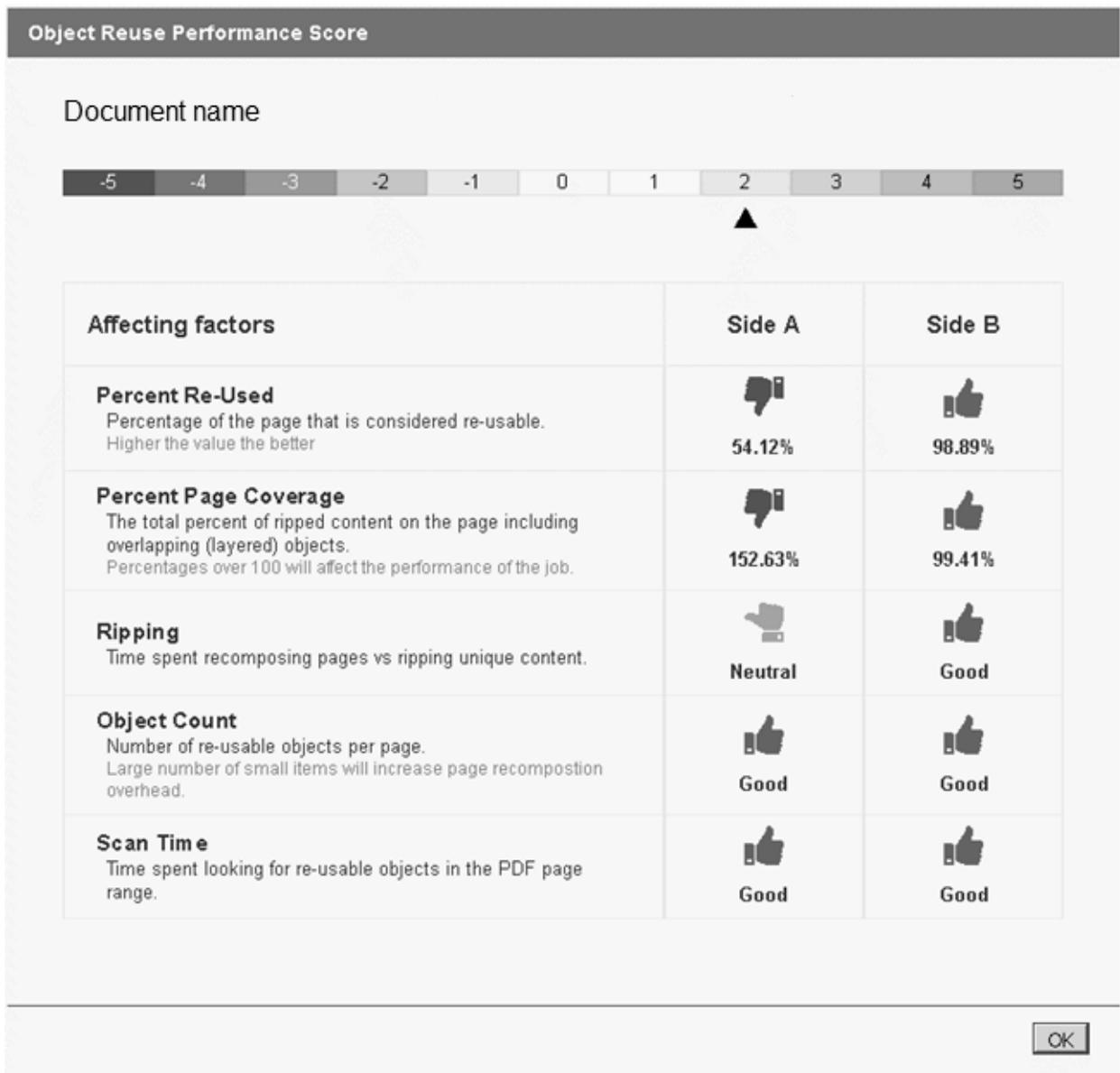


Fig. 2

The object reuse performance score dialog box shows a high-level review of what the ripping system is seeing in the PDF while it is processing the file.

Disclosed by Bryan Crampton, Mahesh Balakrishnan and Balaram Sahu, HP Inc.