Generalization of the BWT Transformation and Inversion Ranks

Ziya Arnavut
Department of Mathematics and Computer Science
SUNY College at Fredonia
Fredonia, NY 14063
Email: arnavut@cs.fredonia.edu

In this paper, we expand the theoretical foundations of LPSA [1] from permutations to multiset permutations (data strings) and give the general theory behind the BWT combinatorially. We then show the information theoretic relationship between interval ranks and canonical sorting permutations. Finally, we explore the relationship between inversion ranks and recency ranks and presents improved compression results over other BSC coders, such as Bzip and Szip.

<table>
<thead>
<tr>
<th>File</th>
<th>Size</th>
<th>BSIC</th>
<th>BSRIC</th>
<th>MBSWC</th>
<th>BZIP2</th>
<th>SZIP-B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bib</td>
<td>111261</td>
<td><strong>1.96</strong></td>
<td>2.03</td>
<td>2.05</td>
<td>1.97</td>
<td><strong>1.96</strong></td>
</tr>
<tr>
<td>Book1</td>
<td>768768</td>
<td><strong>2.22</strong></td>
<td>2.27</td>
<td>2.29</td>
<td>2.42</td>
<td>2.35</td>
</tr>
<tr>
<td>Book2</td>
<td>610856</td>
<td><strong>1.95</strong></td>
<td>1.99</td>
<td>2.02</td>
<td>2.06</td>
<td>2.02</td>
</tr>
<tr>
<td>Geo</td>
<td>102400</td>
<td><strong>4.18</strong></td>
<td>4.47</td>
<td>4.55</td>
<td>4.45</td>
<td>4.28</td>
</tr>
<tr>
<td>News</td>
<td>377109</td>
<td><strong>2.45</strong></td>
<td>2.50</td>
<td>2.55</td>
<td>2.52</td>
<td>2.48</td>
</tr>
<tr>
<td>Obj1</td>
<td>21504</td>
<td>3.88</td>
<td>4.05</td>
<td>4.16</td>
<td>4.01</td>
<td><strong>3.78</strong></td>
</tr>
<tr>
<td>Obj2</td>
<td>246814</td>
<td>2.54</td>
<td>2.56</td>
<td>2.6</td>
<td>2.48</td>
<td><strong>2.47</strong></td>
</tr>
<tr>
<td>Paper1</td>
<td>53161</td>
<td><strong>2.45</strong></td>
<td>2.52</td>
<td>2.59</td>
<td>2.49</td>
<td>2.49</td>
</tr>
<tr>
<td>Paper2</td>
<td>82199</td>
<td><strong>2.36</strong></td>
<td>2.41</td>
<td>2.49</td>
<td>2.44</td>
<td>2.43</td>
</tr>
<tr>
<td>Pic</td>
<td>513216</td>
<td><strong>0.70</strong></td>
<td>0.72</td>
<td>0.74</td>
<td>0.78</td>
<td>0.80</td>
</tr>
<tr>
<td>Progc</td>
<td>39611</td>
<td><strong>2.50</strong></td>
<td>2.58</td>
<td>2.68</td>
<td>2.53</td>
<td><strong>2.50</strong></td>
</tr>
<tr>
<td>Progl</td>
<td>71646</td>
<td>1.74</td>
<td>1.79</td>
<td>1.86</td>
<td>1.74</td>
<td><strong>1.72</strong></td>
</tr>
<tr>
<td>Progp</td>
<td>49379</td>
<td><strong>1.74</strong></td>
<td>1.80</td>
<td>1.85</td>
<td>1.74</td>
<td>1.77</td>
</tr>
<tr>
<td>Trans</td>
<td>93695</td>
<td>1.55</td>
<td>1.60</td>
<td>1.63</td>
<td><strong>1.53</strong></td>
<td>1.54</td>
</tr>
<tr>
<td>Bible</td>
<td>4047392</td>
<td><strong>1.46</strong></td>
<td>1.49</td>
<td>1.50</td>
<td>1.67</td>
<td>1.53</td>
</tr>
<tr>
<td>E.coli</td>
<td>4638690</td>
<td><strong>1.94</strong></td>
<td>1.99</td>
<td><strong>1.94</strong></td>
<td>2.16</td>
<td>2.06</td>
</tr>
<tr>
<td>World192</td>
<td>2473400</td>
<td><strong>1.34</strong></td>
<td>1.37</td>
<td>1.40</td>
<td>1.58</td>
<td>1.40</td>
</tr>
<tr>
<td>Avg.</td>
<td><strong>2.17</strong></td>
<td>2.24</td>
<td>2.29</td>
<td>2.27</td>
<td>2.21</td>
<td></td>
</tr>
<tr>
<td>W. Avg.</td>
<td><strong>1.71</strong></td>
<td>1.75</td>
<td>1.76</td>
<td>1.91</td>
<td>1.80</td>
<td></td>
</tr>
</tbody>
</table>

References

