

UNIVERSIDADE FEDERAL DOS VALES DO JEQUITINHONHA E MUCURI
FACULDADE DE CIÊNCIAS EXATAS
CURSO DE SISTEMAS DE INFORMAÇÃO

**PRAGMASUM: UM SUMARIZADOR AUTOMÁTICO DE TEXTOS BASEADO
EM PERFIL DE USUÁRIO.**

Valdir Junior Cordeiro Rocha

Diamantina
2014

UNIVERSIDADE FEDERAL DOS VALES DO JEQUITINHONHA E MUCURI
FACULDADE DE CIÊNCIAS EXATAS

**PRAGMASUM: UM SUMARIZADOR AUTOMÁTICO DE TEXTOS BASEADO
EM PERFIL DE USUÁRIO.**

Valdir Junior Cordeiro Rocha

Orientador:
Marcus Vinícius Carvalho Guelpeli

Trabalho de Conclusão de Curso apresentado ao
Curso de Sistemas de Informação, como parte
dos requisitos exigidos para a conclusão do
curso.

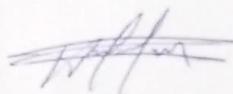
Diamantina
2014

Monografia de projeto final de graduação sob o título "PragmaSUM: Um sumarizador automático de textos baseado em perfil de usuário", defendida por Valdir Júnior Cordeiro Rocha e aprovada em 25 de julho de 2014, em Diamantina, Minas Gerais.

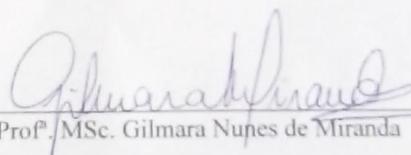
Banca Examinadora:



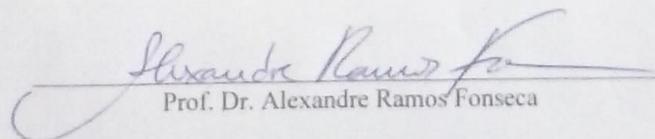
Prof. Dr. Marcus Vinícius Guelpeli
Orientador



Prof. MSc. André Luiz Covre



Prof. MSc. Gilmara Nupes de Miranda



Prof. Dr. Alexandre Ramos Fonseca

*A minha família por todo apoio que me foi dedicado
em mais esta decisão de minha vida.*

AGRADECIMENTOS

Devo agradecer primeiro à minha família, em especial aos meus pais, pois foram eles que me proporcionaram a oportunidade de estar escrevendo isso hoje. Agradeço por todo apoio, suporte e amor que me foi dado em todas as escolhas da minha vida, por toda confiança e paciência em todos esses anos fora de casa, sem dúvida nenhuma, tudo que sou hoje eu devo a vocês.

Às minhas irmãs, meus primos e tio, queridos, por descobrirem na distância o companheirismo e amizade para nos manterem unidos, e proporcionar grandes momentos de alegria.

Ao meu orientador, professor Marcus, por todo tempo dedicado a minha orientação.

Aos amigos do DTI e Biblioteca, por esses quase dois anos de aprendizado, uma grande contribuição para minha vida pessoal e profissional.

À Tássia, pelo companheirismo e paciência com as intermináveis horas de filmes e séries ao meu lado.

A todos os amigos, que de alguma forma, contribuíram para minha formação, em especial ao André, Piolho, Raoni, Luno e Renan sem vocês eu não teria chegado até aqui, bora formar galera.

Enfim, a todos que de alguma forma deixaram uma contribuição para minha caminhada até aqui, o meu muito obrigado.

RESUMO

Este trabalho apresenta o sumarizador automático de textos PragmaSUM, que se baseia na frequência das palavras do texto para valoração das sentenças, a criação de um corpus na língua portuguesa formado por artigos científicos de 10 domínios de conhecimento diferentes e a avaliação dos sumários gerados pelo BLMSumm, o GistSumm e o PragmaSUM com utilização das palavras chave do corpus e sem utilizá-las. O desempenho foi medido pelas métricas *Recall*, *Precision* e *F-Measure* presentes na ferramenta ROUGE e validados com os testes estatísticos ANOVA de Friedman e Coeficiente de concordância de Kendall. Os resultados são animadores obtidos com uso das palavras chave para valoração das sentenças comparados com os sumarizadores analisados.

Palavras chave: PragmaSUM, Sumarizador Automático de Textos, BLMSumm, GistSumm, corpus, ROUGE.

ABSTRACT

This paper presents the automatic summarizer of texts PragmaSUM, based on the frequency of words in the text for the valuation of the sentences, the creation of a corpus in Portuguese composed of 10 scientific articles on different fields of knowledge and the evaluation of summaries generated by BLMSumm the GistSumm and PragmaSUM with the use of key words in the corpus and without using them. Performance was measured by metrics Recall, Precision and F-Measure present in ROUGE tool and validated with the Friedman ANOVA and Kendall coefficient of concordance. The results are encouraging obtained with the use of keywords for valuation of sentences compared to the summarizers analyzed.

Keywords: *PragmaSUM, Automatic summarization of texts, BLMSumm, GistSumm, corpus, ROUGE.*

LISTA DE FIGURAS

Figura 1. Texto Fonte.	8
Figura 2. Sumário automático gerado pelo PragmaSUM com 80% de compressão referente ao texto fonte da figura 1.	9
Figura 3. Esquema simplificado da avaliação de sumários utilizando a ferramenta ROUGE (DELGADO, 2010).	12
Figura 4. Curva de Zipf (Guelpeli, 2012).	17
Figura 5. Curva de Zipf com os cortes de Luhn (Guelpeli (2012)).	18
Figura 6. Seleção das palavras do corte (Guelpeli (2012)).	19
Figura 7. Interface do PragmaSUM – Tela Principal.	21
Figura 8. Interface do PragmaSUM – Tela Sumarização em lote.	23
Figura 9. Algoritmo do PragmaSUM	24
Figura 10. Seleção do Texto Fonte do PragmaSUM.	25
Figura 11. Seleção de palavras do perfil.	26
Figura 12. Sumário gerado com o PragmaSUM.	27
Figura 13. Salvando o sumario gerado com o PragmaSUM.	27
Figura 14. Pasta raiz do corpus.	29
Figura 15. Conteúdo da pasta do domínio Administração.	29
Figura 16. Diagrama do Corpus criado para teste.	30
Figura 17. Arquitetura do GistSumm (Pardo, 2002).	37
Figura 18. Modelo do BLMSumm (Oliveira e Guelpeli, 2011).	38
Figura 19. Comparação entre os resultados de <i>Precision</i> do domínio Administração.	41
Figura 20. Comparação entre os resultados de <i>Precision</i> do domínio Agronomia.	42
Figura 21. Comparação entre os resultados de <i>Precision</i> do domínio Educação Física.	43
Figura 22. Comparação entre os resultados de <i>Precision</i> do domínio Engenharia.	44
Figura 23. Comparação entre os resultados de <i>Precision</i> do domínio Filosofia.	45
Figura 24. Comparação entre os resultados de <i>Precision</i> do domínio Fisioterapia.	46
Figura 25. Comparação entre os resultados de <i>Precision</i> do domínio Geografia.	47
Figura 26. Comparação entre os resultados de <i>Precision</i> do domínio História.	48
Figura 27. Comparação entre os resultados de <i>Precision</i> do domínio Medicina.	49
Figura 28. Comparação entre os resultados de <i>Precision</i> do domínio Psicologia.	50
Figura 29. Comparação de <i>Precision</i> entre todos os domínios com taxa de compressão de 50%.	51
Figura 30. Comparação de <i>Precision</i> entre todos os domínios com taxa de compressão de 70%.	53
Figura 31. Comparação de <i>Precision</i> entre todos os domínios com taxa de compressão de 80%.	54
Figura 32. Comparação de <i>Precision</i> entre todos os domínios com taxa de compressão de 90%.	55
Figura 33. Média acumulada de F-Measure com taxa de compressão de 50% do domínio Administração.	65

Figura 34. Média acumulada de Precision com taxa de compressão de 50% do domínio Administração.....	66
Figura 35. Média acumulada de Recall com taxa de compressão de 50% do domínio Administração.....	66
Figura 36. Média acumulada de F-Measure com taxa de compressão de 70% do domínio Administração.....	67
Figura 37. Média acumulada de Precision com taxa de compressão de 70% do domínio Administração.....	67
Figura 38. Média acumulada de Recall com taxa de compressão de 70% do domínio Administração.....	68
Figura 39. Média acumulada de F-Measure com taxa de compressão de 80% do domínio Administração.....	68
Figura 40. Média acumulada de Precision com taxa de compressão de 80% do domínio Administração.....	69
Figura 41. Média acumulada de Recall com taxa de compressão de 80% do domínio Administração.....	69
Figura 42. Média acumulada de F-Measure com taxa de compressão de 90% do domínio Administração.....	70
Figura 43. Média acumulada de Precision com taxa de compressão de 90% do domínio Administração.....	70
Figura 44. Média acumulada de Recall com taxa de compressão de 90% do domínio Administração.....	71
Figura 45. Comparação dos resultados de F-Measure do domínio Administração.....	72
Figura 46. Comparação dos resultados de Recall do domínio Administração.....	73
Figura 47. Média acumulada de F-Measure com taxa de compressão de 50% do domínio Agronomia.....	74
Figura 48. Média acumulada de Precision com taxa de compressão de 50% do domínio Agronomia.....	74
Figura 49. Média acumulada de Recall com taxa de compressão de 50% do domínio Agronomia.....	75
Figura 50. Média acumulada de F-Measure com taxa de compressão de 70% do domínio Agronomia.....	75
Figura 51. Média acumulada de Precision com taxa de compressão de 70% do domínio Agronomia.....	76
Figura 52. Média acumulada de Recall com taxa de compressão de 70% do domínio Agronomia.....	76
Figura 53. Média acumulada de F-Measure com taxa de compressão de 80% do domínio Agronomia.....	77
Figura 54. Média acumulada de Precision com taxa de compressão de 80% do domínio Agronomia.....	77
Figura 55. Média acumulada de Recall com taxa de compressão de 80% do domínio agronomia.....	78
Figura 56. Média acumulada de F-Measure com taxa de compressão de 90% do domínio Agronomia.....	78

Figura 57. Média acumulada de Precision com taxa de compressão de 90% do domínio Agronomia.....	79
Figura 58. Média acumulada de Recall com taxa de compressão de 90% do domínio Agronomia.....	79
Figura 59. Comparação dos resultados de F-Measure do domínio agronomia.....	80
Figura 60. Comparação dos resultados de Recall do domínio agronomia.....	81
Figura 61. Média acumulada de F-Measure com taxa de compressão de 50% do domínio Educação Física.....	82
Figura 62. Média acumulada de Precision com taxa de compressão de 50% do domínio Educação Física.....	82
Figura 63. Média acumulada de F-Measure com taxa de compressão de 70% do domínio Educação Física.....	83
Figura 64. Média acumulada de Precision com taxa de compressão de 70% do domínio Educação Física.....	83
Figura 65. Média acumulada de Recall com taxa de compressão de 70% do domínio Educação Física.....	84
Figura 66. Média acumulada de F-Measure com taxa de compressão de 80% do domínio Educação Física.....	84
Figura 67. Média acumulada de Precision com taxa de compressão de 80% do domínio Educação Física.....	85
Figura 68. Média acumulada de Recall com taxa de compressão de 80% do domínio Educação Física.....	85
Figura 69. Média acumulada de F-Measure com taxa de compressão de 90% do domínio Educação Física.....	86
Figura 70. Média acumulada de Precision com taxa de compressão de 90% do domínio Educação Física.....	86
Figura 71. Média acumulada de Recall com taxa de compressão de 90% do domínio Educação Física.....	87
Figura 72. Comparação dos resultados de F-Measure do domínio Educação Física.....	88
Figura 73. Comparação dos resultados de Recall do domínio Educação Física.....	89
Figura 74. Média acumulada de F-Measure com taxa de compressão de 50% do domínio Engenharia.....	90
Figura 75. Média acumulada de Precision com taxa de compressão de 50% do domínio Engenharia.....	90
Figura 76. Média acumulada de Recall com taxa de compressão de 50% do domínio Engenharia.....	91
Figura 77. Média acumulada de F-Measure com taxa de compressão de 70% do domínio Engenharia.....	91
Figura 78. Média acumulada de Precision com taxa de compressão de 70% do domínio Engenharia.....	92
Figura 79. Média acumulada de Recall com taxa de compressão de 70% do domínio Engenharia.....	92
Figura 80. Média acumulada de F-Measure com taxa de compressão de 80% do domínio Engenharia.....	93

Figura 81. Média acumulada de Precision com taxa de compressão de 80% do domínio Engenharia.....	93
Figura 82. Média acumulada de Recall com taxa de compressão de 80% do domínio Engenharia.....	94
Figura 83. Média acumulada de F-Measure com taxa de compressão de 90% do domínio Engenharia.....	94
Figura 84. Média acumulada de Precision com taxa de compressão de 90% do domínio Engenharia.....	95
Figura 85. Média acumulada de Recall com taxa de compressão de 90% do domínio Engenharia.....	95
Figura 86. Comparação dos resultados de F-Measure do domínio Engenharia.....	96
Figura 87. Comparação dos resultados de Recall do domínio Engenharia.....	97
Figura 88. Média acumulada de F-Measure com taxa de compressão de 50% do domínio Filosofia.....	98
Figura 89. Média acumulada de Precision com taxa de compressão de 50% do domínio Filosofia.....	98
Figura 90. Média acumulada de Recall com taxa de compressão de 50% do domínio Filosofia.....	98
Figura 91. Média acumulada de F-Measure com taxa de compressão de 70% do domínio Filosofia.....	99
Figura 92. Média acumulada de Precision com taxa de compressão de 70% do domínio Filosofia.....	99
Figura 93. Média acumulada de Recall com taxa de compressão de 70% do domínio Filosofia.....	100
Figura 94. Média acumulada de F-Measure com taxa de compressão de 80% do domínio Filosofia.....	100
Figura 95. Média acumulada de Precision com taxa de compressão de 80% do domínio Filosofia.....	101
Figura 96. Média acumulada de Recall com taxa de compressão de 80% do domínio Filosofia.....	101
Figura 97. Média acumulada de F-Measure com taxa de compressão de 90% do domínio Filosofia.....	102
Figura 98. Média acumulada de Precision com taxa de compressão de 90% do domínio Filosofia.....	102
Figura 99. Média acumulada de Recall com taxa de compressão de 90% do domínio Filosofia.....	103
Figura 100. Comparação dos resultados de F-Measure do domínio Filosofia.....	104
Figura 101. Comparação dos resultados de Recall do domínio Filosofia.....	105
Figura 102. Média acumulada de F-Measure com taxa de compressão de 50% do domínio Fisioterapia.....	106
Figura 103. Média acumulada de Precision com taxa de compressão de 50% do domínio Fisioterapia.....	106
Figura 104. Média acumulada de Recall com taxa de compressão de 50% do domínio Fisioterapia.....	107

Figura 105. Média acumulada de F-Measure com taxa de compressão de 70% do domínio Fisioterapia.	107
Figura 106. Média acumulada de Precision com taxa de compressão de 70% do domínio Fisioterapia.	108
Figura 107. Média acumulada de Recall com taxa de compressão de 70% do domínio Fisioterapia.	108
Figura 108. Média acumulada de F-Measure com taxa de compressão de 80% do domínio Fisioterapia.	109
Figura 109. Média acumulada de Precision com taxa de compressão de 80% do domínio Fisioterapia.	109
Figura 110. Média acumulada de Recall com taxa de compressão de 80% do domínio Fisioterapia.	110
Figura 111. Média acumulada de F-Measure com taxa de compressão de 90% do domínio Fisioterapia.	110
Figura 112. Média acumulada de Precision com taxa de compressão de 90% do domínio Fisioterapia.	111
Figura 113. Média acumulada de Recall com taxa de compressão de 90% do domínio Fisioterapia.	111
Figura 114. Comparação dos resultados de F-Measure do domínio Fisioterapia.	112
Figura 115. Comparação dos resultados de Recall do domínio Fisioterapia.	113
Figura 116. Média acumulada de F-Measure com taxa de compressão de 50% do domínio Geografia.	114
Figura 117. Média acumulada de Precision com taxa de compressão de 50% do domínio Geografia.	114
Figura 118. Média acumulada de Recall com taxa de compressão de 50% do domínio Geografia.	115
Figura 119. Média acumulada de F-Measure com taxa de compressão de 70% do domínio Geografia.	115
Figura 120. Média acumulada de Precision com taxa de compressão de 70% do domínio Geografia.	116
Figura 121. Média acumulada de Recall com taxa de compressão de 70% do domínio Geografia.	116
Figura 122. Média acumulada de F-Measure com taxa de compressão de 80% do domínio Geografia.	117
Figura 123. Média acumulada de Precision com taxa de compressão de 80% do domínio Geografia.	117
Figura 124. Média acumulada de Recall com taxa de compressão de 80% do domínio Geografia.	118
Figura 125. Média acumulada de F-Measure com taxa de compressão de 90% do domínio Geografia.	118
Figura 126. Média acumulada de Precision com taxa de compressão de 90% do domínio Geografia.	119
Figura 127. Média acumulada de Recall com taxa de compressão de 90% do domínio Geografia.	119
Figura 128. Comparação dos resultados de F-Measure do domínio Geografia.	120

Figura 129. Comparação dos resultados de Recall do domínio Geografia.	121
Figura 130. Média acumulada de F-Measure com taxa de compressão de 50% do domínio História.	122
Figura 131. Média acumulada de Precision com taxa de compressão de 50% do domínio História.	122
Figura 132. Média acumulada de Recall com taxa de compressão de 50% do domínio História.	123
Figura 133. Média acumulada de F-Measure com taxa de compressão de 70% do domínio História.	123
Figura 134. Média acumulada de Precision com taxa de compressão de 70% do domínio História.	124
Figura 135. Média acumulada de Recall com taxa de compressão de 70% do domínio História.	124
Figura 136. Média acumulada de F-Measure com taxa de compressão de 80% do domínio História.	125
Figura 137. Média acumulada de Precision com taxa de compressão de 80% do domínio História.	125
Figura 138. Média acumulada de Recall com taxa de compressão de 80% do domínio História.	126
Figura 139. Média acumulada de F-Measure com taxa de compressão de 90% do domínio História.	126
Figura 140. Média acumulada de Precision com taxa de compressão de 90% do domínio História.	127
Figura 141. Média acumulada de Recall com taxa de compressão de 90% do domínio História.	127
Figura 142. Comparação dos resultados de F-Measure do domínio História.	128
Figura 143. Comparação dos resultados de Recall do domínio História.	129
Figura 144. Média acumulada de F-Measure com taxa de compressão de 50% do domínio Medicina.	130
Figura 145. Média acumulada de Precision com taxa de compressão de 50% do domínio Medicina.	130
Figura 146. Média acumulada de Recall com taxa de compressão de 50% do domínio Medicina.	131
Figura 147. Média acumulada de F-Measure com taxa de compressão de 70% do domínio Medicina.	131
Figura 148. Média acumulada de Precision com taxa de compressão de 70% do domínio Medicina.	132
Figura 149. Média acumulada de Recall com taxa de compressão de 70% do domínio Medicina.	132
Figura 150. Média acumulada de F-Measure com taxa de compressão de 80% do domínio Medicina.	133
Figura 151. Média acumulada de Precision com taxa de compressão de 80% do domínio Medicina.	133
Figura 152. Média acumulada de Recall com taxa de compressão de 80% do domínio Medicina.	134

Figura 153. Média acumulada de F-Measure com taxa de compressão de 90% do domínio Medicina.	134
Figura 154. Média acumulada de Precision com taxa de compressão de 90% do domínio Medicina.	135
Figura 155. Média acumulada de Recall com taxa de compressão de 90% do domínio Medicina.	135
Figura 156. Comparação dos resultados de F-Measure do domínio Medicina.	136
Figura 157. Comparação dos resultados de Recall do domínio Medicina.	137
Figura 158. Média acumulada de F-Measure com taxa de compressão de 50% do domínio Psicologia.	138
Figura 159. Média acumulada de Precision com taxa de compressão de 50% do domínio Psicologia.	138
Figura 160. Média acumulada de Recall com taxa de compressão de 50% do domínio Psicologia.	139
Figura 161. Média acumulada de F-Measure com taxa de compressão de 70% do domínio Psicologia.	139
Figura 162. Média acumulada de Precision com taxa de compressão de 70% do domínio Psicologia.	140
Figura 163. Média acumulada de Recall com taxa de compressão de 70% do domínio Psicologia.	140
Figura 164. Média acumulada de F-Measure com taxa de compressão de 80% do domínio Psicologia.	141
Figura 165. Média acumulada de Precision com taxa de compressão de 80% do domínio Psicologia.	141
Figura 166. Média acumulada de Recall com taxa de compressão de 80% do domínio Psicologia.	142
Figura 167. Média acumulada de F-Measure com taxa de compressão de 90% do domínio Psicologia.	142
Figura 168. Média acumulada de Precision com taxa de compressão de 90% do domínio Psicologia.	143
Figura 169. Média acumulada de Recall com taxa de compressão de 90% do domínio Psicologia.	143
Figura 170. Comparação dos resultados de F-Measure do domínio Psicologia.	144
Figura 171. Comparação dos resultados de Recall do domínio Psicologia.	145
Figura 172. Comparação da F-Measure para taxa de compressão de 50%.	146
Figura 173. Comparação da F-Measure para taxa de compressão de 70%.	147
Figura 174. Comparação da F-Measure para taxa de compressão de 80%.	148
Figura 175. Comparação da F-Measure para taxa de compressão de 90%.	149
Figura 176. Diagrama para escolha da técnica teste estatístico a partir do número de amostras (CALLEGARI-JACQUES, 2007).	182

LISTA DE TABELAS

Tabela 1. Estudos relevantes na área de SA.	9
Tabela 2. Estatísticas do corpus criado para testes.	36
Tabela 3. Comparação entre os resultados de <i>Precision</i> do domínio Administração.	41
Tabela 4. Comparação entre os resultados de <i>Precision</i> do domínio Agronomia.	42
Tabela 5. Comparação entre os resultados de <i>Precision</i> do domínio Educação Física.	43
Tabela 6. Comparação entre os resultados de <i>Precision</i> do domínio Engenharia.	44
Tabela 7. Comparação entre os resultados de <i>Precision</i> do domínio Filosofia.	45
Tabela 8. Comparação entre os resultados de <i>Precision</i> do domínio Fisioterapia.	46
Tabela 9. Comparação entre os resultados de <i>Precision</i> do domínio Geografia.	47
Tabela 10. Comparação entre os resultados de <i>Precision</i> do domínio História.	48
Tabela 11. Comparação entre os resultados de <i>Precision</i> do domínio Medicina.	49
Tabela 12. Comparação entre os resultados de <i>Precision</i> do domínio Psicologia.	50
Tabela 13. Comparação de <i>Precision</i> entre todos os domínios com taxa de compressão de 50%.	51
Tabela 14. Comparação de <i>Precision</i> entre todos os domínios com taxa de compressão de 70%.	52
Tabela 15. Comparação de <i>Precision</i> entre todos os domínios com taxa de compressão de 80%.	53
Tabela 16. Comparação de <i>Precision</i> entre todos os domínios com taxa de compressão de 90%.	55
Tabela 17. Comparação dos resultados de F-Measure do domínio Administração.	71
Tabela 18. Comparação dos resultados de Recall do domínio Administração.	72
Tabela 19. Comparação dos resultados de F-Measure do domínio Agronomia.	80
Tabela 20. Comparação dos resultados de Recall do domínio Agronomia.	80
Tabela 21. Comparação dos resultados de F-Measure do domínio Educação Física.	87
Tabela 22. Comparação dos resultados de Recall do domínio Educação Física.	88
Tabela 23. Comparação dos resultados de F-Measure do domínio Engenharia.	95
Tabela 24. Comparação dos resultados de Recall do domínio Engenharia.	96
Tabela 25. Comparação dos resultados de F-Measure do domínio Filosofia.	103
Tabela 26. Comparação dos resultados de Recall do domínio Filosofia.	104
Tabela 27. Comparação dos resultados de F-Measure do domínio Fisioterapia.	112
Tabela 28. Comparação dos resultados de Recall do domínio Fisioterapia.	112
Tabela 29. Comparação dos resultados de F-Measure do domínio Geografia.	120
Tabela 30. Comparação dos resultados de Recall do domínio Geografia.	120
Tabela 31. Comparação dos resultados de F-Measure do domínio História.	127
Tabela 32. Comparação dos resultados de Recall do domínio História.	128
Tabela 33. Comparação dos resultados de F-Measure do domínio Medicina.	136
Tabela 34. Comparação dos resultados de Recall do domínio Medicina.	136
Tabela 35. Comparação dos resultados de F-Measure do domínio Psicologia.	143
Tabela 36. Comparação dos resultados de Recall do domínio Psicologia.	144
Tabela 37. Comparação da F-Measure para taxa de compressão de 50%.	145
Tabela 38. Comparação da F-Measure para taxa de compressão de 70%.	147

Tabela 39. Comparação da F-Measure entre os domínios médico e jornalístico para taxa de compressão de 80%.	148
Tabela 40. Comparação da F-Measure para taxa de compressão de 90%.	149
Tabela 41. Teste estatístico da métrica F-Measure – Domínio Administração.	151
Tabela 42. Teste estatístico da métrica Precision – Domínio Administração.	152
Tabela 43. Teste estatístico da métrica Recall – Domínio Administração.	153
Tabela 44. Teste estatístico da métrica F-Measure – Domínio Agronomia.	154
Tabela 45. Teste estatístico da métrica Precision – Domínio Agronomia.	155
Tabela 46. Teste estatístico da métrica Recall – Domínio Agronomia.	156
Tabela 47. Teste estatístico da métrica F-Measure – Domínio Educação Física.	157
Tabela 48. Teste estatístico da métrica Precision – Domínio Educação Física.	158
Tabela 49. Teste estatístico da métrica Recall – Domínio Educação Física.	159
Tabela 50. Teste estatístico da métrica F-Measure – Domínio Engenharia.	160
Tabela 51. Teste estatístico da métrica Precision – Domínio Engenharia.	161
Tabela 52. Teste estatístico da métrica Recall – Domínio Engenharia.	162
Tabela 53. Teste estatístico da métrica F-Measure – Domínio Filosofia.	163
Tabela 54. Teste estatístico da métrica Precision – Domínio Filosofia.	164
Tabela 55. Teste estatístico da métrica Recall – Domínio Filosofia.	165
Tabela 56. Teste estatístico da métrica F-Measure – Domínio Fisioterapia.	166
Tabela 57. Teste estatístico da métrica Precision – Domínio Fisioterapia.	167
Tabela 58. Teste estatístico da métrica Recall – Domínio Fisioterapia.	168
Tabela 59. Teste estatístico da métrica F-Measure – Domínio Geografia.	169
Tabela 60. Teste estatístico da métrica Precision – Domínio Geografia.	170
Tabela 61. Teste estatístico da métrica Recall – Domínio Geografia.	171
Tabela 62. Teste estatístico da métrica F-Measure – Domínio História.	172
Tabela 63. Teste estatístico da métrica Precision – Domínio História.	173
Tabela 64. Teste estatístico da métrica Recall – Domínio História.	174
Tabela 65. Teste estatístico da métrica F-Measure – Domínio Medicina.	175
Tabela 66. Teste estatístico da métrica Precision – Domínio Medicina.	176
Tabela 67. Teste estatístico da métrica Recall – Domínio Medicina.	177
Tabela 68. Teste estatístico da métrica F-Measure – Domínio Psicologia.	178
Tabela 69. Teste estatístico da métrica Precision – Domínio Psicologia.	179
Tabela 70. Teste estatístico da métrica Recall – Domínio Psicologia.	180

LISTA DE SIGLAS

ANOVA – *ANalysis Of VAriance*

LC – Linguística de *Corpus*

PLN – Processamento de Linguagem Natural

ROUGE – *Recall-Oriented Understudy for Gisting Evaluation*

SA – Sumarização Automática

TAC – *Text Analysis Conference*

SUMÁRIO

1.	INTRODUÇÃO	1
1.1	Problema/Motivação	2
1.2	Hipótese.....	2
1.3	Objetivo.....	2
1.4	Contribuições	2
1.5	Estrutura da Proposta	3
2.	FUNDAMENTAÇÃO TEÓRICA	5
2.1	Corpus	5
2.2	Sumarização.....	6
2.2.1	Sumarização Automática.....	6
2.3	Métricas	10
2.4	Rouge.....	11
2.5	Testes Estatísticos.....	12
2.5.1	ANOVA de Friedman.....	13
2.5.2	Coefficiente de concordância de Kendall	14
3.	PRAGMASUM	17
3.1	Seleção de palavras no PragmaSUM.....	17
3.2	Pré-processamento no PragmaSUM.....	20
3.3	Processamento no PragmaSUM	22
3.3.1	Algoritmo	23
3.4 .	Exemplos de sumarização no PragmaSUM.....	25
4.	METODOLOGIA	28
4.1	Corpora	28
4.2	Sumarizadores	36
4.2.1	GistSumm	37
4.2.2	BLMSumm	38
4.3	Processo de Sumarização Automática.....	39
5.	RESULTADOS	40
5.1.1	Domínio Administração	40
5.1.2	Domínio Agronomia.....	41
5.1.3	Domínio Educação Física.....	42
5.1.4	Domínio Engenharia.....	43
5.1.5	Domínio Filosofia.....	44
5.1.6	Domínio Fisioterapia	45
5.1.7	Domínio Geografia.....	46
5.1.8	Domínio História	47
5.1.9	Domínio Medicina.....	48

5.1.91 Domínio Psicologia	49
5.2 Comparação entre os Domínios.....	50
5.2.1 Taxa de compressão 50%	50
5.2.2 Taxa de compressão 70%	52
5.2.3 Taxa de compressão 80%	53
5.2.4 Taxa de compressão 90%	54
5.3 Análise dos testes estatísticos	56
5.4 Análise-Hipótese	56
5.5 Discussão	57
6. CONCLUSÃO	59
6.1 Contribuições.....	60
6.2 Limitações	60
6.3 Trabalhos Futuros	61
REFERÊNCIAS	62
APÊNDICES	65
Apêndice A – Tabelas e gráficos com os resultados do f-measure, precision e recall	65
Apêndice B – Tabelas com os resultados dos testes estatísticos do f-measure, precision e recall.....	151
ANEXO	181
ANEXO A	182

1. INTRODUÇÃO

Atualmente, são muitas as informações textuais disponibilizadas na internet, chegando de várias maneiras e com diferentes finalidades, o que torna impossível assimilar todas. Selecionar as que melhor correspondem aos interesses do público facilita o processamento e a recuperação dessas informações (GUELPELI, 2012).

O grande fluxo de informações geradas atualmente torna impossível a leitura de todos os textos encontrados, já que a capacidade de uma pessoa é limitada, além do grande tempo e esforço necessários para isso. Uma alternativa é a apresentação condensada dessas informações, tornando mais fácil e rápida a análise do conteúdo e a decisão da leitura completa do texto.

Segundo Oliveira (2014), uma das áreas de pesquisa que se dedica a isto é a da Sumarização Automática (SA). A SA é uma subárea de estudo de Processamento de Linguagem Natural (PLN) e propõe a geração de resumos de maneira automatizada com o objetivo de reduzir o volume de informações sem perder a qualidade original do texto fonte. O sistema que realiza esta tarefa é chamado na literatura de sumarizador automático ou sumarizador.

Oliveira (2014) ressalta que, atualmente, embora seja possível encontrar diversos sumarizadores automáticos, seja na literatura ou comerciais, a maioria deles é voltada para o idioma inglês.

Tendo em vista as limitações encontradas na área de SA, este trabalho apresenta a criação de um sumarizador automático, o PragmaSUM, que é independente do idioma e do domínio de conhecimento presentes no texto fonte, e é adaptável ao perfil do usuário.

O PragmaSUM utiliza um novo método de corte para sumarização, apresentado por Guelpeleli (2012) no modelo de clusterização de textos Cassiopeia.

Foi construído um corpus em português, realizados testes de avaliação de desempenho de sumarizadores automáticos e aplicadas métricas *Recall*, *Precision* e *F-Measure* utilizado a ferramenta ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*) (LIN e HOVY, 2003).

Com este trabalho espera-se poder contribuir, principalmente, apresentando uma nova forma de SA, com um modelo que melhore a precisão dos sumários gerados. Espera-se, ainda, considerando as circunstâncias propostas por este trabalho, viabilizar estudos sobre a importância da escolha das palavras chave em artigos científicos.

1.1 Problema/Motivação

Os sumarizadores automáticos existentes na literatura atualmente não apresentam métodos de personificação do sumário.

1.2 Hipótese

Um sumário automático que consiga sumarizar um texto fonte independente de seu idioma e de seu domínio de conhecimento, além de respeitar a taxa de compressão definida, e que se adapte ao perfil do leitor, poderia gerar sumários mais personificados e com o tamanho desejado pelo usuário, mantendo ou melhorando o desempenho que os sumarizadores existentes conseguem alcançar.

1.3 Objetivo

O principal objetivo deste trabalho é apresentar o sumário automático de texto que se adapte ao perfil do usuário. Para isso foi desenvolvido o PragmaSUM, um sumário automático que utiliza palavras escolhidas pelo usuário para personificar o sumário de acordo com suas preferências.

1.4 Contribuições

A partir dos experimentos e das análises dos resultados dos sumários gerados pelo PragmaSUM, apresentados neste trabalho, podem-se destacar algumas contribuições importantes para a área de sumarização automática de textos:

- Criar um modelo que melhore a precisão do sumário.
- Criar sumários mais personificados, de acordo com preferências do usuário;
- Viabilizar um estudo sobre a importância das escolhas das palavras chaves em artigo científico;

- Possibilitar uma estrutura algorítmica para uso do aprendizado de máquina;
- Viabilizar um estudo da importância da palavra chave para o processo de indexação de artigos científicos.

1.5 Estrutura da Proposta

O trabalho desenvolvido e a estrutura desta monografia estão organizados nos seguintes capítulos que serão brevemente descritos:

Capítulo 2 – Fundamentação Teórica

Neste capítulo será apresentado o embasamento teórico sobre o qual está fundamentada a pesquisa. Serão definidos o conceito de *corpus*, de sumarização, tipos de sumários, tipos de abordagens e um histórico sobre SA. Serão apresentadas, ainda, as métricas que serão utilizadas nas avaliações dos sumarizadores automáticos e, finalmente, os testes estatísticos ANOVA (*ANalysis Of VAriance*) de Friedman e Concordância de Kendall.

Capítulo 3 – PragmaSUM

Este capítulo descreve a construção do PragmaSUM. Primeiramente serão apresentados os critérios utilizados para sumarização, logo em seguida, o seu algoritmo. Posteriormente, seguirá uma descrição sobre a interação com a interface do PragmaSUM para a sumarização de textos.

Capítulo 4 – Metodologia

Este capítulo descreve a metodologia seguida na construção do corpus para os testes realizados. Primeiramente serão apresentados os critérios utilizados na coleta e organização dos textos e, logo em seguida, as estatísticas que o compõem. Posteriormente, seguirá uma descrição sobre os sumarizadores automáticos que foram avaliados, bem como sobre a ferramenta ROUGE utilizada na avaliação. Por fim, virão descritos os procedimentos aplicados nos experimentos de sumarização dos textos.

Capítulo 5 – Resultados

O capítulo traz os resultados das simulações e as comparações feitas entre cada sumariador. Mostra tabelas e gráficos nos quais se podem perceber as alterações causadas pelas mudanças de domínio, taxas de compressão e utilização de palavras chave na sumarização com o PragmaSUM. O capítulo traz também a comprovação da hipótese e análise através dos testes estatísticos ANOVA de Friedman e Concordância de Kendall.

Capítulo 6 – Conclusão

Este capítulo mostra as contribuições e as limitações deste trabalho e traz sugestões para serem aplicadas em trabalhos futuros.

2. FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta a fundamentação teórica para o entendimento dos conceitos básicos da pesquisa aqui apresentada. Primeiramente será apresentado o conceito de Corpus e Sumarização e depois as métricas e os testes estatísticos utilizados para avaliação dos resultados deste trabalho.

2.1 Corpus

Segundo Oliveira (2014), a palavra *corpus* (plural, *corpora*) vem do latim *corpo*, conjunto de textos, que em Linguística de *Corpus* (LC) determina uma coleção de textos selecionados e organizados segundo critérios para servir à pesquisa científica.

Para um melhor entendimento sobre *corpus* é necessário o entendimento do termo linguístico. A LC é uma área que estuda a língua por meio da observação de grandes quantidades de dados linguísticos reais, isto é, textos falados ou escritos provenientes da comunicação no mundo real, ou seja, da língua em uso, com o auxílio de ferramentas computacionais (SARDINHA, 2004).

De acordo com SARDINHA (2004) o *corpus* pode ser definido como uma coletânea de porções de linguagem que são selecionadas e organizadas de acordo com critérios linguísticos explícitos, a fim de serem usadas como uma amostra de linguagem, ou seja, são conjuntos de textos agrupados com um critério previamente escolhido usado para o estudo ou pesquisa envolvendo linguística, no nosso caso, na sumarização automática de textos.

Segundo Fernandes e Guelpeli (2014), Aluísio e Almeida (2006) apresentam diversas definições de *corpus* e as agrupam sob duas perspectivas diferentes: da Linguística e da Linguística de *Corpus*. A principal diferença entre as duas é relativa ao formato do *corpus*. Para a Linguística, o *corpus* pode ser composto por documentos impressos, enquanto que para Linguística de *Corpus*, o *corpus* deve, obrigatoriamente, estar em formato eletrônico.

Tomando por base esta última perspectiva, Humblé (2001) ressalta que um *corpus* pode ser compreendido como uma quantidade grande de textos armazenados no

computador e que são acessados com programas próprios de pesquisa, em que os textos que fazem parte do *corpus* podem ser variados, indo de jornais a romances, ou se concentrar num determinado gênero, como só periódicos ou só revistas, por exemplo.

Segundo Fernandes e Guelpeli (2014), a criação de *corpora* consiste em gerar conjuntos de dados linguísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Estes dados podem ser obtidos de diversas maneiras dependendo do que se quer trabalhar, sendo as mais comuns: escanear livros, obter textos já digitalizados, e textos oriundos da internet.

2.2 Sumarização

A sumarização, em geral, é uma atividade bastante comum. Quando se narra um evento a uma pessoa, costuma-se fazer um resumo do que aconteceu, e não uma narração completa e detalhada. Inconscientemente, as pessoas estão sempre sumarizando. Muito frequentes também são os sumários escritos, como, por exemplo, notícias em jornais, artigos de revistas, resumo de textos científicos, entre muitos outros (MARTINS *et al*, 2001).

Por definição, sumários são textos reduzidos, que transmitem as ideias principais e mais relevantes de um texto original, de forma clara e objetiva, sem perda da informatividade (PARDO, 2007).

Sumários podem ser classificados como indicativos, informativos ou críticos, segundo Pardo (2002): sumários indicativos apenas listam ou indicam o assunto principal dos textos-fonte; os informativos são autocontidos, isto é, possuem toda a informação essencial dos textos fonte, dispensando a leitura destes; os críticos avaliam ou apenas comentam o conteúdo de suas fontes.

2.2.1 Sumarização Automática

Segundo Pardo (2002), a sumarização automática de textos tem se tornado uma área proeminente devido à crescente demanda por informação no menor tempo possível. Com o advento da internet, onde as pessoas se vêem em um mar de informação em

constante expansão e atualização, um grande interesse acadêmico, comercial e governamental surgiu por essa área.

A sumarização automática é o processo de seleção das informações mais importantes de um texto, aqui chamado de texto fonte, para produzir uma versão mais curta do mesmo, visando determinado usuário e tarefa (PARDO e RINO, 2002).

A Figura 1 mostra um texto¹, aqui chamado de texto fonte, extraído do portal G1 e a Figura 2 mostra seu respectivo sumário gerado pelo sumarizador PragmaSUM. A sumarização foi feita com uma redução de 80% do tamanho em relação ao texto original.

1 Em SP, cartório terá de informar ao Detran sobre venda de veículo, G1, 25 de junho de 2014, Carros. Disponível em: <<http://g1.globo.com/carros/noticia/2014/06/em-sp-cartorio-tera-de-informar-ao-detransobre-venda-de-veiculo.html>>. Acesso em 25 jun. 2014.

Quem vender ou comprar um carro emplacado no estado de São Paulo se verá livre de uma das burocracias do processo: a de ter que ir até o Departamento Estadual de Trânsito (Detran) para informar a venda do carro.

A partir do próximo dia 23, será obrigação do cartório avisar a Secretaria da Fazenda sobre compra, venda ou transferência de propriedade de veículos, quando o proprietário for reconhecer firma para validação do recibo de compra e venda que existe no verso do Certificado de Registro do Veículo (CRV).

A norma foi estabelecida pelo Decreto nº 60.489 do governo de São Paulo.

Caberá aos cartórios enviar cópia digitalizada (frente e verso) do Certificado de Registro do Veículo (CRV) preenchido e com firmas reconhecidas por autenticidade, conforme determinado pela legislação de trânsito, à secretaria, com assinatura digital.

Não será permitido ao cartório cobrar nada além do valor de reconhecimento de firma e da cópia autenticada do CRV.

O notário terá também a opção de enviar as informações e a cópia digitalizada do CRV por lote, no prazo de até 72 horas. Se não cumprir a nova obrigação, estará sujeito a multa.

Se o antigo dono do veículo e o novo proprietário reconhecerem firma simultaneamente, bastará uma única transmissão dos dados.

Quem vende e quem compra

Atualmente, é dado um prazo de 30 dias a quem compra um carro para fazer a transferência, a partir da data da compra que consta recibo de compra e venda do CRV, e a obrigação de validar esse processo é de quem compra o carro.

Continuará sendo obrigação do novo proprietário preencher o recibo de compra e venda, que é encontrado no verso do documento do veículo, e reconhecer firma dele e do antigo dono. O Detran-SP alerta que não basta o antigo dono assinar o CRV e entregá-lo ao novo proprietário: é importante que ele se certifique de que a transferência do documento foi feita.

Os contribuintes poderão obter informações sobre a efetivação da comunicação de venda do veículo na área de serviços eletrônicos do Detran, no endereço www.detran.sp.gov.br.

Figura 1. Texto Fonte.

A partir do próximo dia 23, será obrigação do cartório avisar a Secretaria da Fazenda sobre compra, venda ou transferência de propriedade de veículos, quando o proprietário for reconhecer firma para validação do recibo de compra e venda que existe no verso do Certificado de Registro do Veículo (CRV).

Quem vende e quem compra

Atualmente, é dado um prazo de 30 dias a quem compra um carro para fazer a transferência, a partir da data da compra que consta recibo de compra e venda do CRV, e a obrigação de validar esse processo é de quem compra o carro.

Continuará sendo obrigação do novo proprietário preencher o recibo de compra e venda, que é encontrado no verso do documento do veículo, e reconhecer firma dele e do antigo dono.

Figura 2. Sumário automático gerado pelo PragmaSUM com 80% de compressão referente ao texto fonte da figura 1.

A sumarização automática vem sendo formalizada desde 1950. A Tabela 1 demonstra cronologicamente as informações sobre os estudos mais relevantes da área de sumarização automática apresentados por Guelpeli (2012).

Tabela 1. Estudos relevantes na área de SA.

Ano	Autor /Tipo de estudo
1958	Luhn apresenta o método das palavras-chaves
1958	Baxendale versa sobre a importância da primeira e da última frase do texto original.
1969	Edmundson aborda a escolha computacional das frases com o maior potencial de transmitir significância do texto original.
1975	Pollock e Zamora reforçam relevância da restrição do domínio.
1987	Hutchins classifica os sumários em indicativos, informativos e de críticas
1993	Paice e Jones hierarquizam os papéis semânticos em cada frase.
1993	Maybury propõe o uso da abordagem híbrida.
1997	Hovy e Lin fazem uso do conhecimento simbólico e de técnicas estatísticas para sumarização.
1999	Sparck afirma que a taxonomia criada por Hutchins (1987) é um fator determinante para estabelecer sua aplicabilidade e criar uma avaliação consistente deste processo.

Segundo Martins *et al* (2002), a sumarização automática vem sendo objeto de estudo desde os primórdios da computação. Já no final da década de 50 começaram a surgir alguns métodos estatísticos para extrair as sentenças principais de um texto, sendo o método das palavras-chave (Luhn, 1958) o mais significativo até então. Entretanto os métodos estatísticos, que fazem uso de técnicas de sumarização superficial, geravam extratos que apresentavam inúmeros problemas, como de coesão e coerência, razão pela qual a área ficou praticamente estagnada, voltando a ser objeto de interesse com o advento da internet, o que gerou grande quantidade de documentos *on-line* e a necessidade de processar informações em larga escala.

As pesquisas trouxeram avanços com o uso de duas abordagens principais: a superficial e a profunda, as quais caracterizam métodos distintos de sumarização automática. A superficial utiliza métodos estatísticos para retirar as sentenças de um texto fonte e utilizá-las em um sumário, e a profunda utiliza técnicas de linguística para formá-lo.

Segundo Oliveira e Guelpeli (2011) o maior desafio da área de SA é gerar um sumário que não perca sua informatividade. Sendo assim a avaliação de SA é muito subjetiva, pois depende fundamentalmente da avaliação humana e não existe um consenso entre os avaliadores humanos frente aos resultados obtidos com os sumarizadores automáticos.

2.3 Métricas

Para avaliar os sumários gerados serão utilizadas as métricas estatísticas *Recall* (cobertura), *Precision* (precisão) e *F-Measure* (média harmônica). Na avaliação são comparados os resumos retirados dos artigos científicos com sumários automáticos. Segundo OLIVEIRA (2014) essas métricas são definidas a seguir:

- *Recall* (R): indica o quanto do sumário manual permanece no sumário automático. É calculado através da fórmula (1):

$$R = \frac{N_s SA \in SM}{N_s SA} \quad (1)$$

Onde $N_s SA$ é o número de sentenças do sumário automático e SM é o número de sentenças do sumário manual.

- *Precision* (P): indica o quanto do sumário automático coincide com o sumário manual. É calculado através da fórmula (2):

$$P = \frac{N_s SA \in SM}{SM} \quad (2)$$

- *F-Measure* (FM): indica a média harmônica entre a cobertura e a precisão, e assim, é calculada através da fórmula (3) (ZACARIAS *et al.*, 2012; GUELPELI *et al.*, 2008):

$$FM = \frac{P * R}{(P + R) * 2} \quad (3)$$

O resultado do cálculo ocupa um intervalo entre 0 e 1 e quanto mais próximo de 1 estiver, mais o sumário automático se aproximará do sumário humano comparado (GUELPELI, 2012; LUCHI e RIBEIRO, 2011). Há na literatura uma ferramenta bastante conceituada que utiliza estas métricas, ela é chamada ROUGE (*Recall Oriented Understudy for Gisting Evaluation*) (LIN e HOVY, 2003).

2.4 Rouge

Segundo Oliveira (2014), trata-se de um pacote de avaliação automática de sumários que compara a qualidade dos resumos gerados pelos sumarizadores automáticos com aqueles feitos por seres humanos. Esta ferramenta é adotada em conferências

internacionais dedicadas ao tema, como a TAC (*Text Analysis Conference*), realizada anualmente nos Estados Unidos da América, patrocinada pelo sistema de defesa americano. O uso de avaliação automatizada se justifica pela grande quantidade de textos a serem analisados e pelo alto custo que essa teria se fosse feita por especialistas. A ROUGE, para avaliar, considera a quantidade de *n-gramas*, isto é, o conjunto de palavras em sequência, que têm em comum os sumários gerados automaticamente e os sumários manuais. As *n-gramas* ocupam um intervalo que está entre 0 e 1 e quanto mais próximo de 1 estiver o resultado do cálculo, mais o sumário automático se assemelha ao sumário humano comparado (LUCCHI e RIBEIRO, 2011). A Figura 3 mostra uma estrutura simplificada do processo de avaliação com a ROUGE.

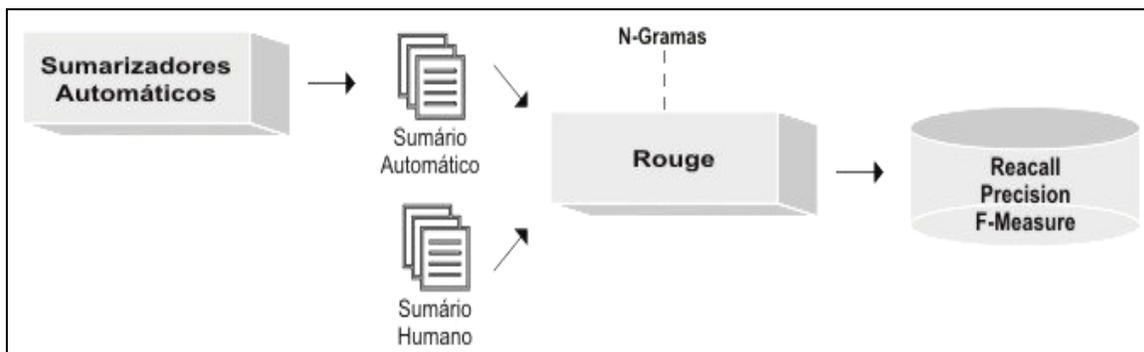


Figura 3. Esquema simplificado da avaliação de sumários utilizando a ferramenta ROUGE (DELGADO, 2010).

2.5 Testes Estatísticos

Neste trabalho a escolha dos testes foi feita de acordo com o Anexo A que traz um diagrama proposto por Callegari-Jacques (2007) e também são usados na maior conferência internacional da área TAC. O diagrama indica que os testes mais adequados para validar esta pesquisa são o ANOVA de Friedman e o coeficiente de Concordância de Kendall. Isto porque os dados aqui examinados são ordinais, as amostras analisadas seguem uma distribuição anormal, sendo, portanto, necessário usar testes não paramétricos, e, ainda, as amostras são independentes e possuem *k* variáveis.

2.5.1 ANOVA de Friedman

O ANOVA de Friedman é um teste não paramétrico que serve para comparar os resultados obtidos por três ou mais amostras relacionadas, numa distribuição bivariada (CALLEGARI-JACQUES, 2007). O teste não utiliza os dados da amostra diretamente, mas sim, a ordem ocupada por eles. O ANOVA ordena os resultados para cada um dos casos e depois calcula a média das ordens para cada uma das amostras.

Segundo Guelpeli (2012), para calcular ordens para cada amostra, a estatística de teste ANOVA de Friedman ordena as k observações da menor para a maior, de forma separada, em cada um dos b blocos, e atribui os ranks $\{1, 2, \dots, k\}$ para cada bloco da tabela de observações (CALLEGARI e JACQUES, 2007).

Assim, a posição esperada de qualquer observação sob H_0 é $(k + 1) / 2$. Sendo $r(X_{ij})$ o rank da observação, X_{ij} define a soma de todos os ranks da coluna j (ou seja, de cada tratamento) mostrado na **Equação 4**:

$$R_j = \sum_{i=1}^b r(X_{ij}), 1 \leq j \leq k \quad (4)$$

Se H_0 é verdadeira, o valor esperado de R_j é $E(R_j) = b(k+1) / 2$. Dessa forma, a estatística é mostrada na **Equação 5**:

$$\sum_{i=1}^b (R_j - \frac{b(k+1)}{2})^2 \quad (5)$$

A estatística do teste de Friedman será dada pela **Equação 6**:

$$SFr = \frac{12b}{k(k+1)} \sum_{i=1}^k (\frac{R_j}{b} - \frac{k+1}{2})^2 = \left[\frac{12}{bk(k+1)} \sum_{j=1}^k R_j^2 \right] - 3b(k+1) \quad (6)$$

Se $F_j(t) = F(t+ t_j)$ é a função de distribuição do tratamento j , com $j = 1, 2, \dots, k$, no teste de Friedman estar-se-á interessado em testar a hipótese $H_0: t_1 = t_2 = \dots = t_k$ contra a hipótese alternativa de que t_1, t_2, \dots, t_k não são todas iguais. Neste caso, em nível de significância α , rejeita-se a hipótese H_0 se $SFr \geq SFr(\alpha)$. Caso contrário, não se rejeita a hipótese nula, em que a constante $SFr(\alpha)$ é escolhida de modo que a probabilidade de erro do tipo I seja igual α .

A aproximação para amostras grandes, em que H_0 , a estatística SFr tem, quando n tende ao infinito, uma distribuição qui-quadrado X^2 com $k-1$ graus de liberdade. Nesse caso, utilizando a aproximação qui-quadrado, rejeita-se H_0 se $SFr \geq X_{k-1,\alpha}^2$, caso contrário não se rejeita H_0 , onde $X_{k-1,\alpha}^2$ é tal que $P\text{-valor} = P[X_{k-1}^2 \geq X_{k-1,\alpha}^2] = \alpha$.

No caso de observações repetidas entre as k observações de um mesmo bloco, uma modificação para a estatística SFr é necessária. Nesse caso, substitui-se SFr pela **Equação 7**:

$$SFr' = \frac{12 \sum_{j=1}^k R_j^2 - 3b^2k(k+1)^2}{bk(k+1) - \left[\frac{1}{k-1} \right] \sum_{i=1}^n \left\{ \left(\sum_{j=1}^{g_i} t_{i,j}^3 \right) - k \right\}} \quad (7)$$

Onde g_i denota o número de grupos de observações repetidas no i -ésimo bloco e $t_{i,j}$ é o tamanho do j -ésimo grupo de observações repetidas no i -ésimo bloco. Em particular, se não há observações repetidas entre as observações no i -ésimo bloco, então $g_i = k$ e $t_{i,j} = 1$ para cada $j = 1, \dots, k$. Se em todos os blocos não existem observações repetidas, então SFr' se reduz a SFr . Sendo assim, o p-valor é calculado desta forma $P[X_{k-1}^2 \geq EFr' H_0]$.

2.5.2 Coeficiente de concordância de Kendall

O teste de coeficiente de concordância de Kendall tem como objetivo normalizar o ANOVA de Friedman. É um teste não paramétrico que gera uma avaliação de concordância, com *ranks* estabelecidos nos experimentos, e assim, mede a diferença entre a probabilidade das classificações estarem na mesma ordem ou em ordens diferentes. O resultado ocupa um intervalo entre 0 e 1: quanto mais próximo de 1 estiver, maior é a concordância, e quanto mais próximo de 0 menor é a concordância (CALLEGARI-JACQUES, 2007).

Segundo Guelpeli (2012), o coeficiente de concordância de Kendall é uma medida da relação entre vários conjuntos de postos de N objetos ou indivíduos.

Quando se tem k conjuntos de postos, pode-se determinar a associação entre eles utilizando-se o coeficiente de concordância de Kendall.

O coeficiente de concordância de Kendall expressa a associação simultânea (relacionamento) entre as séries r de rankings (por exemplo, os casos de amostras correlacionadas). Basicamente, esse coeficiente mede a diferença entre a probabilidade de as classificações estarem na mesma ordem e a de estarem em ordens diferentes (CALLEGARI-JACQUES, 2007).

Assim, o cálculo de $M_{Kendall}$ é mais simples e $M_{Kendall}$ tem uma relação linear com o valor médio de r , relativo a todos os grupos, denotando $M_{rKendall}$, o valor médio dos coeficientes de correlação por postos de Kendall, mostrado na **Equação 8**:

$$M_{rKendall} = \frac{k * M_{Kendall} - 1}{k - 1} \quad (8)$$

Outro processo consiste em imaginar como se apresentariam dados se não houvesse concordância alguma entre os conjuntos de postos e, em seguida, como se apresentariam se houvesse concordância perfeita.

O coeficiente de concordância seria, então, um índice de divergência entre a concordância efetiva, acusada pelos dados, e a concordância máxima possível (perfeita). De modo que $M_{Kendall}$ é um coeficiente desta natureza.

Verificou-se que o grau de concordância entre os K julgamentos é refletido pelo grau de variância entre as N somas de postos $M_{Kendall}$, coeficiente de concordância, que é uma função desse grau de variância mostrada na **Equação 9**:

$$M_{Kendall} = \frac{S}{\frac{1}{12} k^2 (N^3 - N)} \quad (9)$$

Onde:

S = soma dos quadrados dos desvios observados a contar da média dos R_j , isto é

$$S = \sum (R_j - \sum \frac{R_j}{n})^2 .$$

k = número de conjuntos de postos;

N = número de entidades (objetos ou indivíduos a que se atribuem postos);

$\frac{1}{12}k^2(N^3 - N)$ = valor máximo possível da soma dos quadrados dos desvios, isto é, S que

ocorreria o caso de concordância perfeita entre os k conjuntos de postos.

O efeito dos empates foi reduzir o valor de $M_{Kendall}$. Se a proporção de empates é pequena, então o efeito é desprezível.

Se a proporção de empates é grande, pode introduzir-se uma correção, que aumentará ligeiramente o valor de $M_{Kendall}$ em relação ao que se apresenta sem correção. O elemento corretivo é mostrado na **Equação 10**:

$$E_c = \frac{\sum (c^3 - c)}{12} \quad (10)$$

Onde:

c = número de observações num grupo empatadas em relação a um dado posto.

A **Equação 11** ficaria introduzindo o elemento E_c :

$$M_{Kendall} = \frac{S}{\frac{1}{12}k^2(N^3 - N) - \sum E_c} \quad (11)$$

3. PRAGMASUM

Este capítulo apresenta o PragmaSUM e as técnicas utilizadas na sumarização de textos. Primeiramente será apresentado como acontece sua seleção de palavras, em seguida, o seu processamento, assim como seu algoritmo e um pouco do seu código fonte, e por fim, será apresentado uma breve demonstração da interação com sua interface.

3.1 Seleção de palavras no PragmaSUM

Segundo Guelpeli (2012), a Curva de Zipf (Figura 4) é uma distribuição estatística específica, que se encontra em raros fenômenos estocásticos. Acontece que um deles é a distribuição da frequência da ocorrência de palavras num texto, em que nas ordenadas Y , se tem o valor dessa frequência, e nas abscissas X , o valor da posição de ordenação relativa dessa palavra, em termos da sua frequência em relação ao das outras palavras do texto. Para a Curva Zipf de uma dada amostra específica, tem-se $Y \cdot X = k$, em que k será uma constante específica para essa amostra.

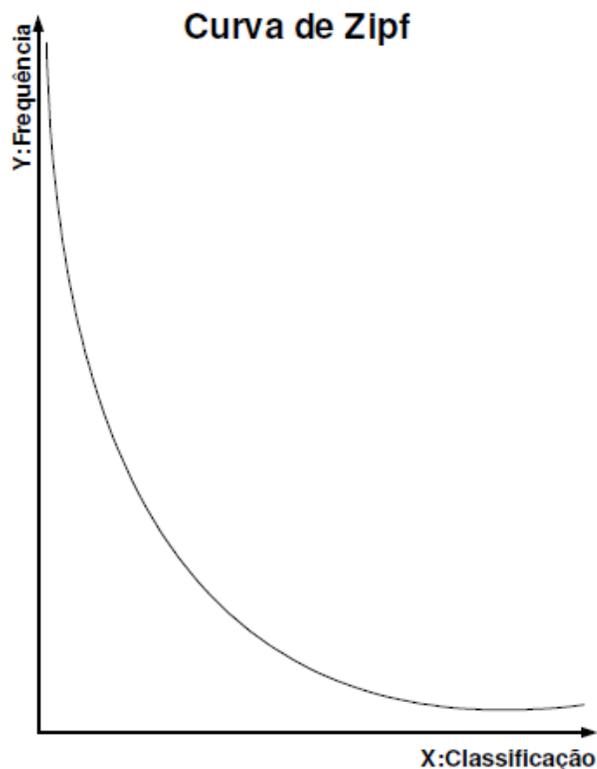


Figura 4. Curva de Zipf (Guelpeli, 2012).

Segundo Guelpeli (2012), Luhn propôs que na Figura 4, se pode definir um limite superior e um inferior de corte denominado limiares de corte de Luhn (Figura 5). Com isso, Luhn propôs uma técnica para encontrar termos relevantes, assumindo que os mais significativos para discriminar o conteúdo do documento estão em um pico imaginário, posicionado no meio dos dois pontos de corte, de acordo com a Figura 5. O primeiro corte de Luhn, conhecido como corte superior, tem por finalidade retirar as *stopwords*, ou seja, as palavras mais frequentes de um texto, o segundo corte serve para diminuir o número de palavras muito específicas, encontradas apenas uma única vez nos documentos, e fazem com que na representação matricial, contribuam para um número grande de dados esparsos. Com isso gera-se três áreas distintas, Na área I, encontram-se as informações triviais ou básicas, com maior frequência; na área II, as informações interessantes e, na área III, os ruídos.

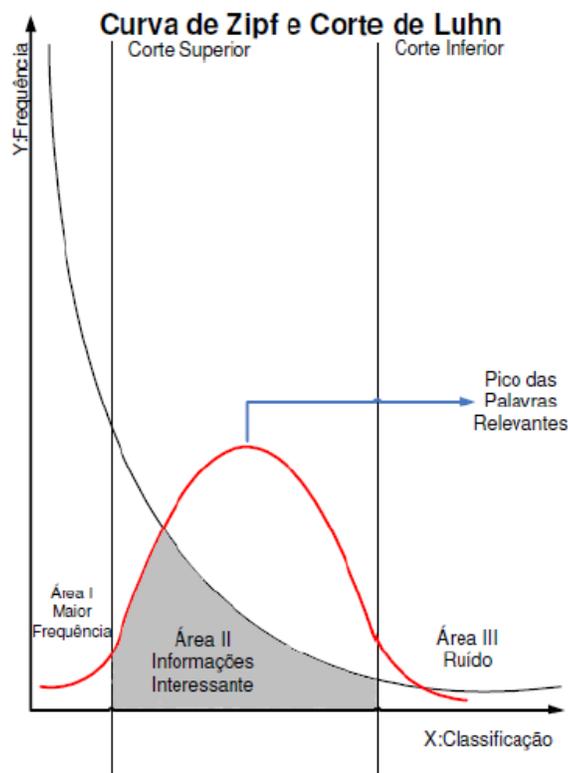


Figura 5. Curva de Zipf com os cortes de Luhn (Guelpeli (2012)).

Levando em consideração essas informações, Guelpeli (2012) propõe em seu modelo de clusterização de textos Cassiopeia uma técnica para redução da alta dimensionalidade e dos dados esparsos. Segundo Guelpeli (2012) tendo como base os pesos das palavras, obtidos na frequência relativa, é calculada a média sobre o total de palavras no documento. Nessa etapa, o modelo usa a truncagem, ou seja, um tamanho máximo de 50 posições para os vetores de palavras, realizando um corte que representa a frequência média das palavras obtidas com os cálculos e, em seguida, realiza a organização dos vetores de palavras (Figura 6). Desse modo elimina-se a maior parte das *stopwords*, que tem maior frequência e as palavras que não tem tanta relevância no texto, e utiliza-se somente as palavras contidas no pico das palavras relevantes para valorar as sentenças do texto.

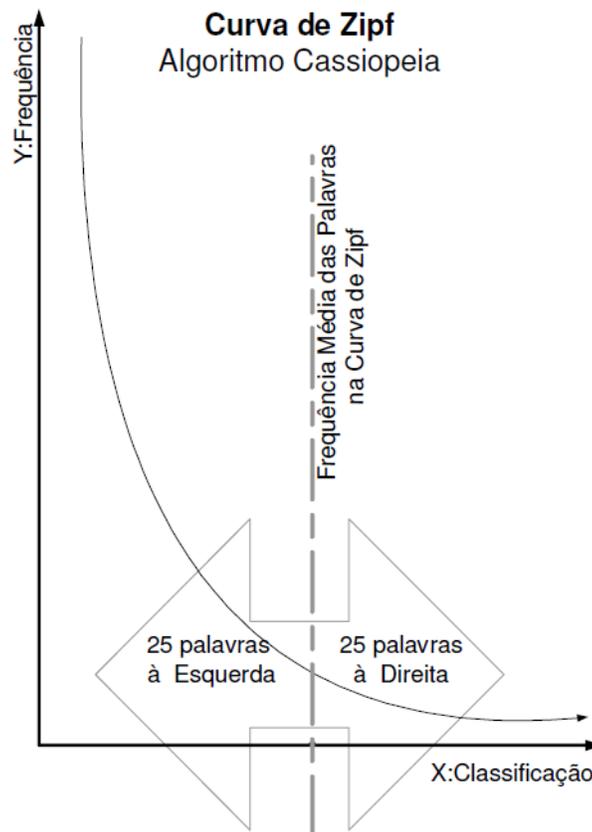


Figura 6. Seleção das palavras do corte (Guelpeli (2012)).

Para isso o PragmaSUM calcula a média da frequência das palavras do texto fonte e escolhe a palavra que tem a frequência mais aproximada da média e seleciona as 24 palavras acima e 25 palavras abaixo desta média, fazendo um corte com 50 palavras que serão usadas na valoração das sentenças do texto.

A taxa de compressão do sumário pode ser escolhida pelo usuário, e é calculada de acordo com o número total de sentenças presentes no texto fonte. A equação 12, a seguir, apresenta o cálculo utilizado pelo PragmaSUM.

$$SS = \frac{ST * TC}{100} \quad (12)$$

Onde:

SS é o número de sentenças do sumário, ST é o número de sentenças do Texto-fonte e TC é a porcentagem do texto que vai estar presente no sumário, dado pela diferença entre 100 e a taxa de compressão escolhida. Se SS não for um número inteiro, é selecionado o inteiro mais próximo.

Para personificar o sumário poderão ser utilizadas 5 palavras que são escolhidas pelo usuário do PragmaSUM, elas são valoradas com 6 para a primeira palavra, 5 para a segunda e assim por diante até a quinta palavra escolhida, com valor 2. Essas palavras podem ser escolhidas pela lista feita pelo corte ou inseridas pelo usuário, desse modo as sentenças que contenham essas palavras tem valores maiores e maior chance de aparecerem no sumário gerado.

A utilização destas palavras para a valoração das sentenças do texto fonte é importante para a personificação do sumário obtido, podendo gerar sumários mais precisos, de acordo com o perfil do usuário, o Capítulo 5 apresenta os resultados da precisão obtida com o uso destas palavras.

3.2 Pré-processamento no PragmaSUM

No pré-processamento ocorre a limpeza dos textos, a preparação para o processo computacional, mas a principal preocupação é a redução do número de palavras, não apenas para viabilizar a questão computacional, mas também para obter a informatividade

das palavras mantidas, ou seja, proporcionar um ganho qualitativo e quantitativo para o processamento (Guelpeli, 2012).

É nesta etapa que as palavras são selecionadas. Através do cálculo da média da frequência de todas as palavras do texto, é selecionado a palavra que tem a frequência mais próxima da média, assim definindo o ponto de corte. Assim é selecionada a palavra onde será feito o corte, a partir dessa palavra seleciona mais 24 palavras com frequência acima e 25 palavras com frequência abaixo deste corte.

É também nesta etapa que são definidas as palavras do perfil do usuário para valoração das sentenças, estas palavras podem ser inseridas pelo usuário ou escolhidas da lista de palavras geradas do corte do algoritmo Cassiopeia. A Figura 7 apresenta a interface da tela principal do PragmaSUM, pode-se notar onde será apresentado as palavras do corte e as 5 palavras escolhidas para compor o perfil.

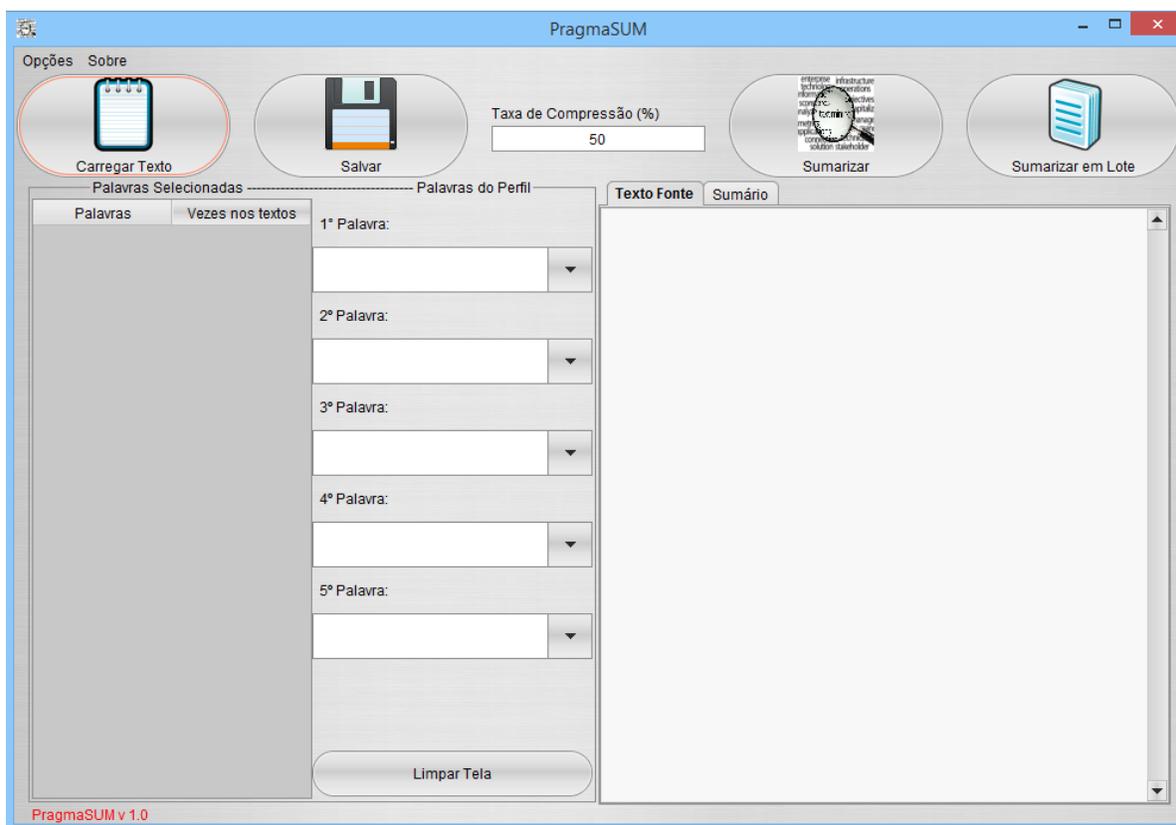


Figura 7. Interface do PragmaSUM – Tela Principal.

3.3 Processamento no PragmaSUM

O PragmaSUM é um sumariador automático de textos independente do idioma e domínio, ele foi desenvolvido com a linguagem de programação Java, utilizando do ambiente de desenvolvimento Eclipse para o sistema operacional Windows. Por ser desenvolvido em Java, o PragmaSUM pode ser executado em qualquer sistema operacional que tenha a máquina virtual Java(JVM) instalada.

O sumariador apresenta duas formas de sumariização, a primeira, em que é sumariado apenas um texto fonte, com a opção de utilizar ou não de palavras para personificação do sumário de acordo com as preferências do usuário, e a segunda, que sumaria textos em lote, ou seja, grandes quantidades de textos de uma só vez. Para sumariização em lote é preciso passar como referência a pasta onde está contida os textos fonte e a pasta onde se deseja salvar os sumários. Estes sumários serão salvos de acordo com a nomenclatura que o texto fonte estiver com a adição do sufixo SUM, assim um arquivo que estiver com o nome Artigo.txt terá seu sumário com o nome ArtigoSUM.txt.

Na sumariização em lote, as palavras do corte são selecionadas de acordo com todas as palavras do grupo de textos analisados, e, assim como na sumariização normal, essas palavras podem ser escolhidas ou inseridas pelo usuário para compor as palavras do perfil.

A Figura 8 apresenta a tela de sumariização em lote. Para a criação da interface gráfica e todos os seus componentes foram utilizadas as bibliotecas awt e swing do Java.

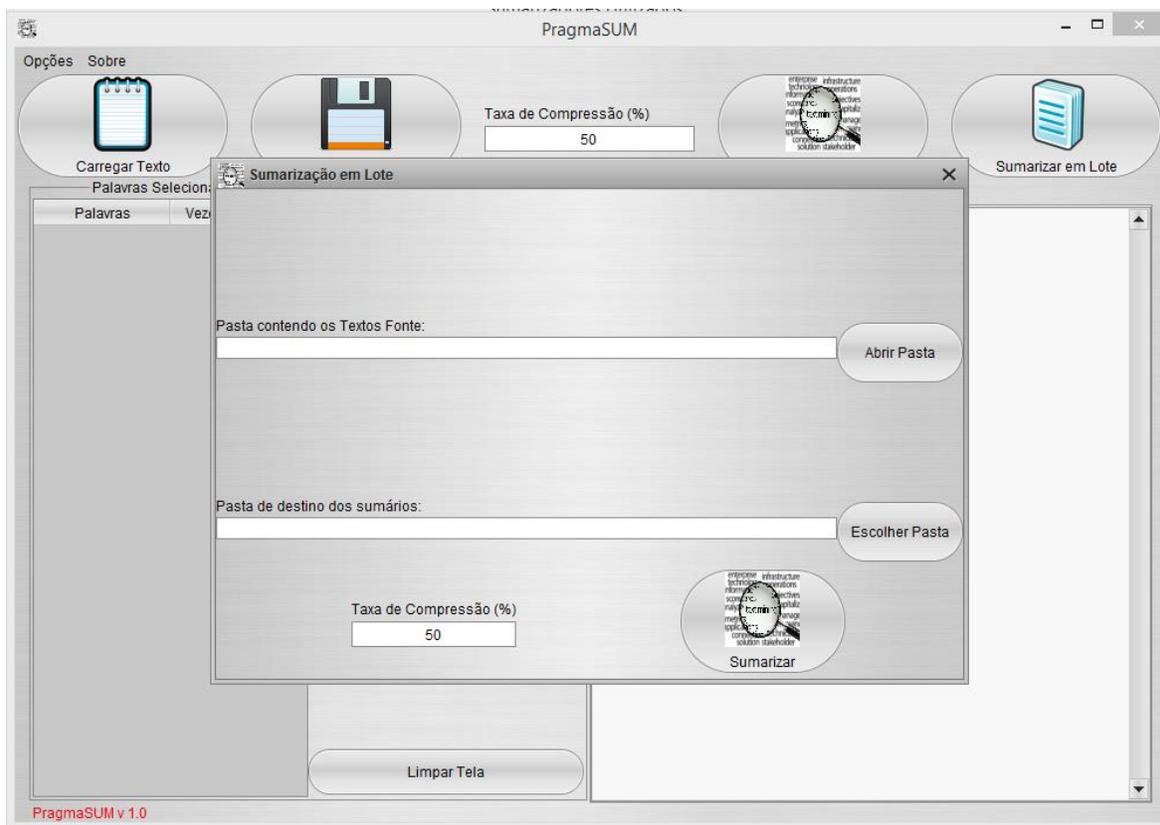


Figura 8. Interface do PragmaSUM – Tela Sumarização em lote.

3.3.1 Algoritmo

Para um entendimento detalhado do processo de sumarização utilizado no PragmaSUM, nesta seção será descrito passo a passo o seu algoritmo. Os procedimentos serão assim representados: {} e dividido em duas fases, a do pré-processamento e a da sumarização, como mostrado na Figura 9.

Pré-Processamento

1. Selecionar Texto Fonte para sumarização.
2. Selecionar as palavras do perfil do usuário.
3. Selecionar a taxa de compressão do texto fonte.

Sumarização

4. {Separar todas as sentenças do texto em uma lista}
5. {Separar todas as palavras de cada sentença}
6. Calcular a média da frequência de todas as palavras do texto: somando as frequências das palavras e dividindo pelo número total de palavras do texto fonte;
7. Ordenar as palavras em ordem decrescente de frequência (da maior para a menor);
8. Encontrar a primeira palavra cuja frequência mais próxima à média;
9. Marcar esta palavra e escolher, incluindo-a, mais as 24 anteriores (esquerda);
10. Marcar esta palavra e escolher as 25 posteriores (direita);
11. Montar uma lista com as 50 palavras escolhidas.
12. {Valorar as sentenças} 1 ponto para cada aparição de uma palavra da lista criada.
13. Se forem escolhidas palavras para o perfil
 - a. {Valorar as sentenças com perfil} 6 pontos para a primeira palavra do perfil, 5 para a segunda, 4 para a terceira, 3 para a quarta e 2 pontos para a quinta palavra escolhida.
14. Ordenar a lista de sentenças de acordo com o valor obtido, do maior para o menor.
15. {Calcular o número de sentenças do sumário} dado pelo número de sentenças do sumário multiplicado pela diferença entre 100 e a taxa de compressão escolhida e o resultado dividido por 100. Caso o valor obtido não for um número real, escolher o inteiro mais próximo.
16. Selecionar o número de sentenças definidas anteriormente, do maior valor para o menor e montar em uma lista.
17. Ordenar a lista de sentenças da forma em que elas aparecem no texto.
18. {Gerar o sumário} com as sentenças desta lista.

Figura 9. Algoritmo do PragmaSUM

3.4 . Exemplos de sumarização no PragmaSUM

Esta seção apresenta um exemplo da interação do usuário com o PragmaSUM na sumarização de textos. A Figura 10 apresenta a tela de seleção do texto fonte, que aparece depois de clicado no botão Carregar Texto.

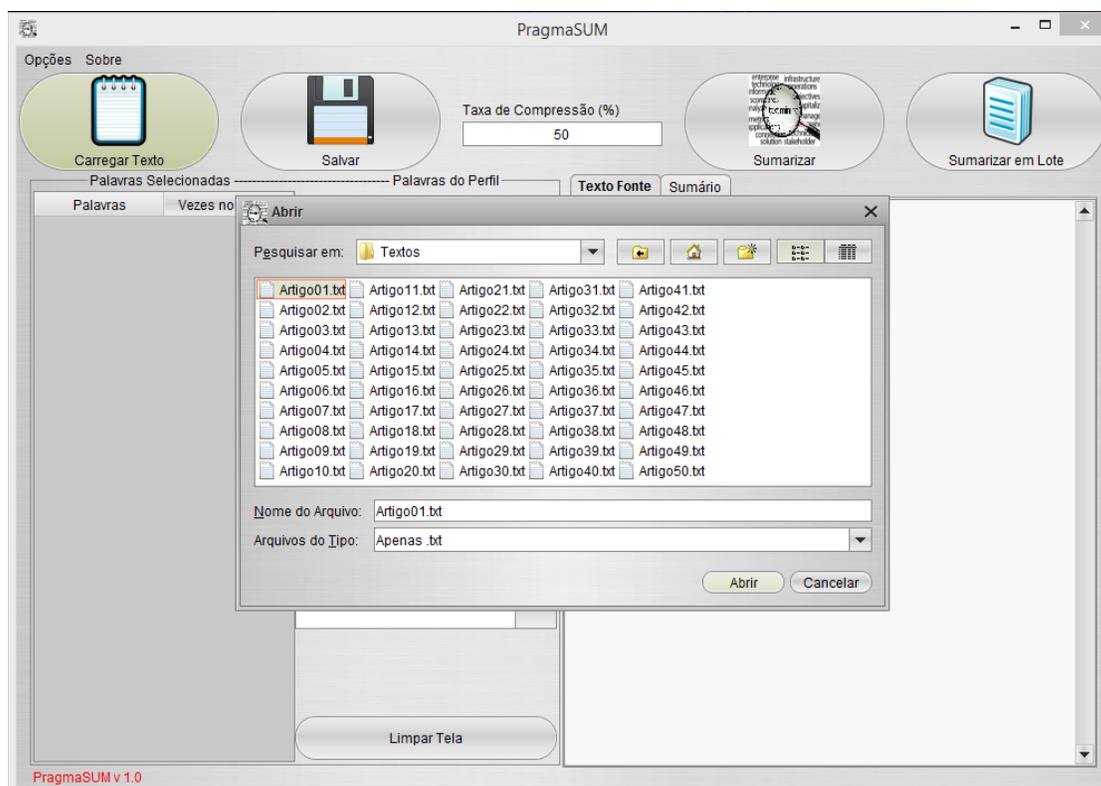


Figura 10. Seleção do Texto Fonte do PragmaSUM.

Como apresentado na Figura 11, depois de selecionado o texto, as palavras selecionadas aparecem e há a opção de escolher 5 delas ou inserir manualmente as palavras para sumarização personalizada. Há, também, a opção de selecionar a taxa de compressão.

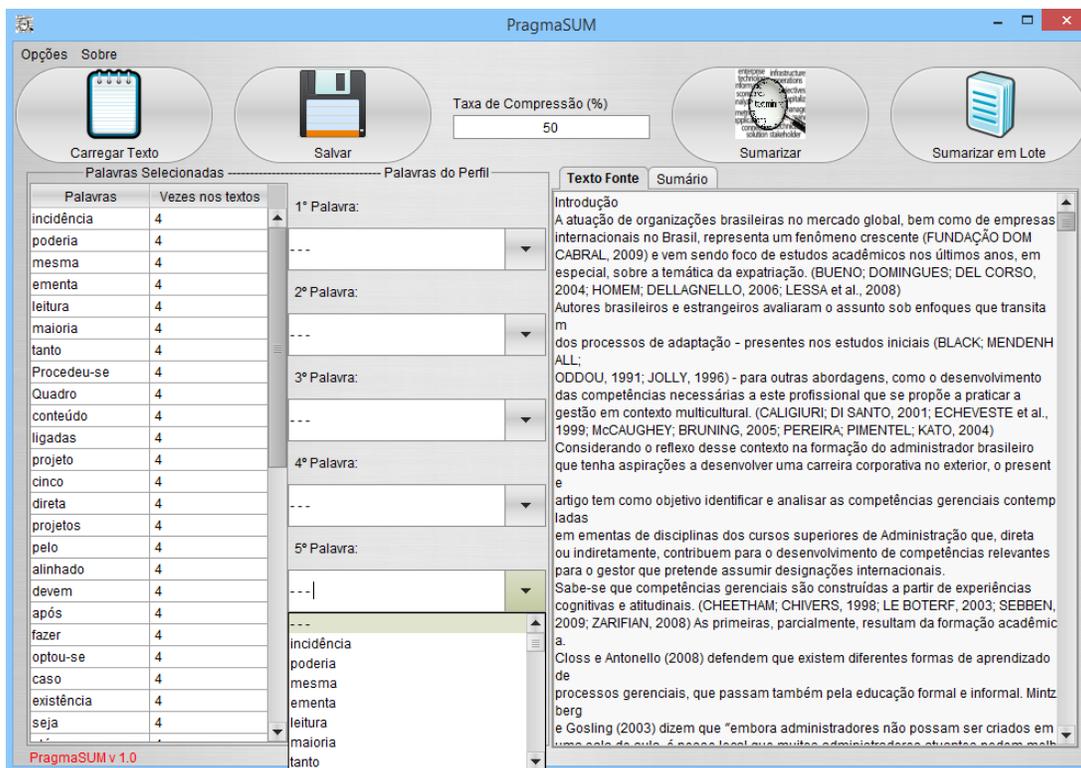


Figura 11. Seleção de palavras do perfil.

A Figura 12 apresenta o sumário gerado pelo PragmaSUM depois de escolhidas as palavras do perfil e a taxa de compressão de 90%. Depois do sumário gerado a opção de salvar um sumario em arquivo de texto pode ser usada, como apresentado na Figura 13.



Figura 12. Sumário gerado com o PragmaSUM.

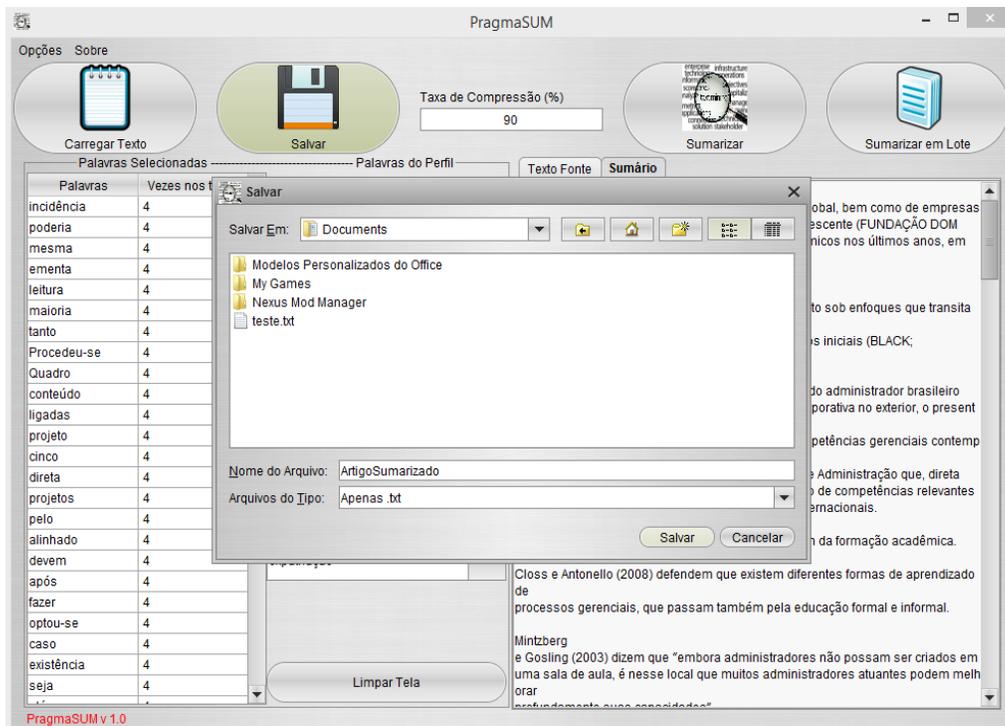


Figura 13. Salvando o sumario gerado com o PragmaSUM.

4. METODOLOGIA

Este capítulo descreve a metodologia utilizada para os testes realizados com os sumarizadores, a construção do corpus de teste, os domínios escolhidos, sua organização e as fontes de onde foram extraídos. Serão apresentados os critérios utilizados na coleta e organização dos dados textuais e as estatísticas que compõem o corpus. Posteriormente, segue uma descrição sobre os sumarizadores automáticos que foram avaliados e, por fim, são apresentados os experimentos aplicados na avaliação destes sumarizadores.

4.1 Corpora

O *corpus* construído neste trabalho é constituído de artigos científicos em português e possui textos em dez domínios de conhecimento diferentes. A escolha por artigos científicos se deu por eles conterem resumos dos seus respectivos autores, que foram usados na avaliação pelo ROUGE, e também pelas palavras chave de cada artigo que foram utilizadas na sumarização de textos com o PragmaSUM. Todos os artigos científicos foram retirados da base de dados *Web of Science* em um total de 500 artigos separados em 10 domínios de conhecimento diferentes, cada um contendo 50 artigos.

Na fase de captura e manipulação foram retiradas dos textos as bibliografias, as tabelas, as notas de rodapé, as figuras, os números de página e os gráficos, restando somente o corpo do texto. Os arquivos foram convertidos dos formatos originais para o formato *txt* a fim de possibilitar o processamento computacional (GUELPELI, 2012).

Cada artigo foi separado em três arquivos de texto, o primeiro contém o resumo disponibilizado pelo autor, o segundo, as palavras chave, e o terceiro o corpo do texto. Os arquivos foram organizados como mostra a Figura 14, dentro de pastas da seguinte maneira: na pasta raiz *Corpus* para teste existem as 10 pastas de cada domínio de conhecimento.

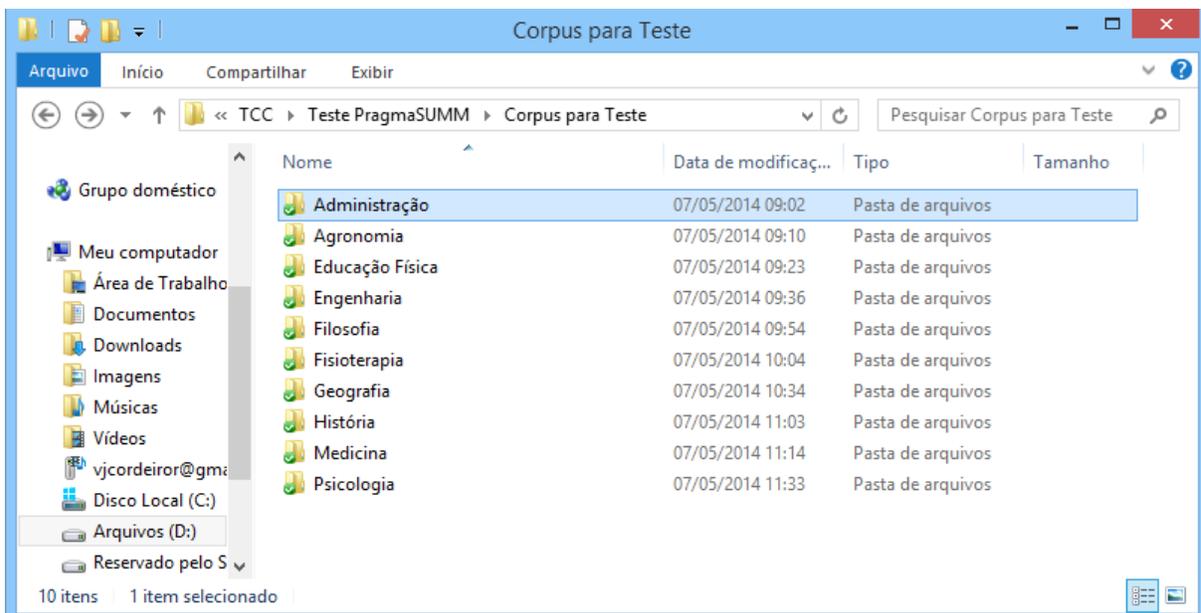


Figura 14. Pasta raiz do corpus.

A Figura 15 apresenta a organização de cada domínio, que são quatro diretórios organizados da seguinte forma: Textos que trazem os arquivos em .txt para serem processados, Resumos que contém arquivos com os resumos produzidos pelos autores de cada artigo, Palavras-Chave que contém as palavras chave escolhidas pelos autores de cada artigo e Artigos Originais que contém os artigos originais em formato pdf.

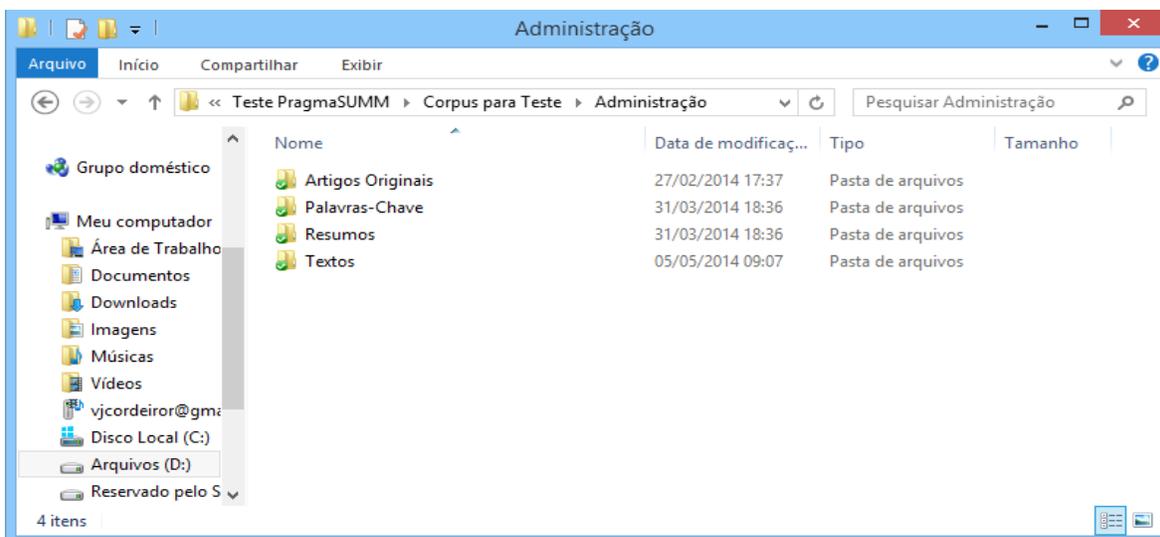


Figura 15. Conteúdo da pasta do domínio Administração.

O *corpus* possui 2000 arquivos de texto, sendo 200 de cada domínio de conhecimento. Cada domínio possui 50 artigos originais, 50 textos fonte, 50 resumos e 50 arquivos contendo as palavras chave. A Figura 16 mostra o diagrama do *Corpus* criado para os testes.

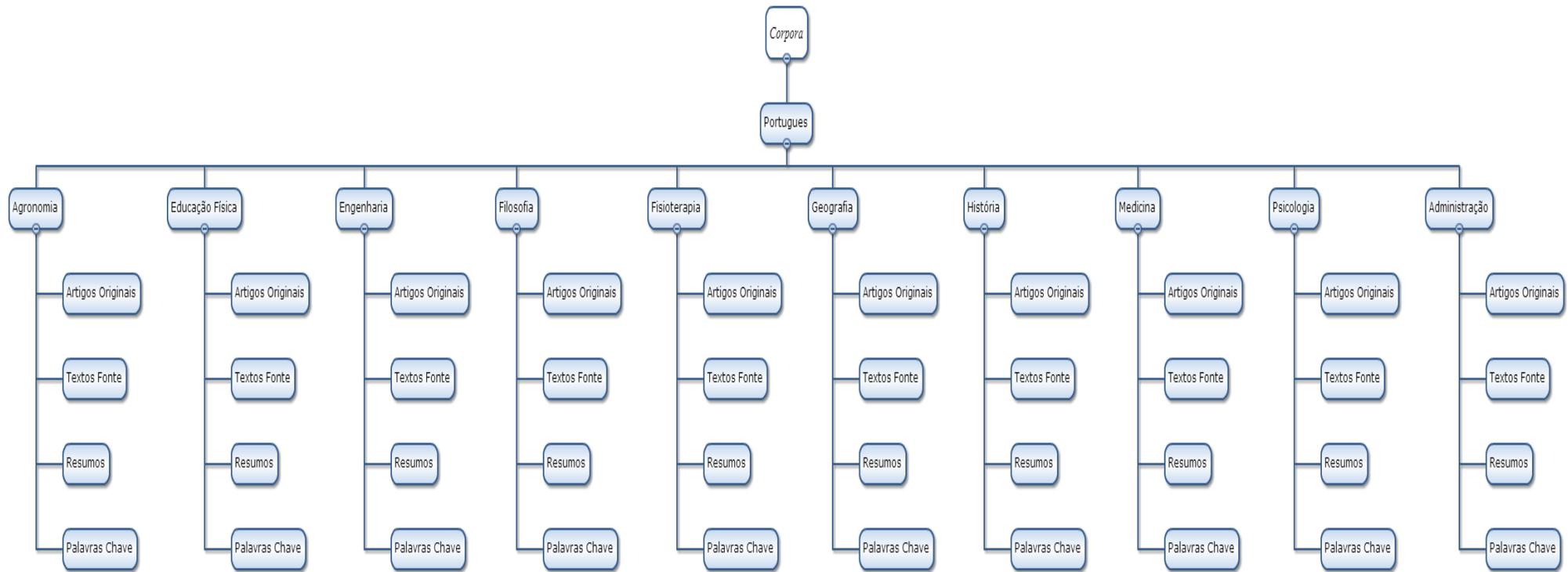


Figura 16. Diagrama do Corpus criado para teste.

Para gerar as estatísticas que compõem o *Corpus* de teste fez-se uso da versão *free* do software *FineCount 2.6*. A Tabela 2 traz as estatísticas dos textos dos 10 domínios de conhecimento que formam o *corpus*. Cada domínio contém 50 textos.

Tabela 2. Estatísticas do corpus criado para testes.

Arquivos	Caracteres	Caracteres e espaços	Palavras	Palavras e numerais	Sentenças	Média de palavras por texto
<i>Administração</i>	2132914	2634837	424805	435561	25362	8496
<i>Agronomia</i>	684889	873911	137806	146621	13384	2756
<i>Educação Física</i>	1469394	1841347	298644	306003	23392	5972
<i>Engenharia</i>	1563482	2008393	313304	322307	19588	6266
<i>Filosofia</i>	1761247	2202861	371484	377061	26260	7429
<i>Fisioterapia</i>	769856	924618	153446	160104	21129	3068
<i>Geografia</i>	1482445	1775764	298354	306404	31658	5967
<i>História</i>	2057161	2469530	425457	432658	40889	8509
<i>Medicina</i>	849235	1016013	168232	176688	19571	3364
<i>Psicologia</i>	1529982	1830251	305705	311233	32760	6114,1
Total	14300605	17577525	2897237	2974640	253993	57941
Desvio Padrão	511808,2	630245,2	105794,1	105857,2	7947,024	2116,011
Média Geral	1430060,5	1757752,5	289723,7	297464	25399,3	5794,1

4.2 Sumarizadores

Todos os sumarizadores utilizados são automáticos e da literatura. A escolha deles se deu por conta deles serem capazes de sumarizarem na língua portuguesa e possibilitarem adotar uma taxa de compressão do texto entre 50% a 90%.

4.2.1 GistSumm

O GistSumm – GIST SUMMArizer (Pardo, 2002) é um sumarizador que se baseia na ideia principal do texto através do qual é possível se identificar a sentença que melhor representa a ideia principal do texto, denominada sentença-gist pelo próprio Pardo(2002).

O GistSumm utiliza uma abordagem superficial, ou seja, utiliza métodos estatísticos para identificar a sentença-gist ou a sentença que mais se aproxime dela. A Figura 17 apresenta a arquitetura do GistSumm.

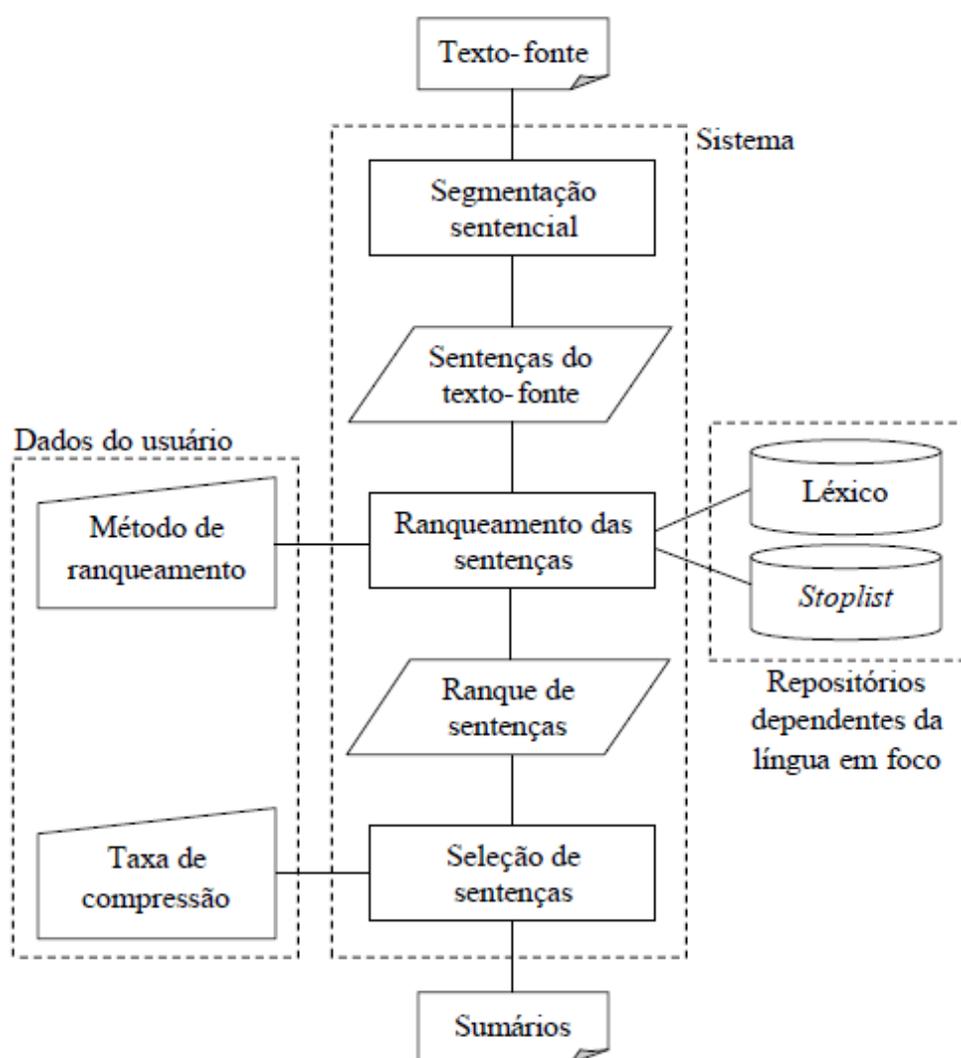


Figura 17. Arquitetura do GistSumm (Pardo, 2002).

A nova versão do GistSumm (Pardo, 2005) foi utilizada neste trabalho. Foram implementadas novas extensões na nova versão, como a sumarização multidocumento, onde vários textos fonte dão origem a apenas um sumário, outros métodos de ranqueamento de sentenças adicionados, mas que não foram utilizadas nas sumarizações realizadas neste trabalho. Segundo Guelpeli (2012), afirmam Pardo *et al.* (2006) que o GistSumm foi ampliado, também para tratamento na utilização de texto científico. Esse foi um dos principais motivos para adotar a última versão, já que todos os textos utilizados neste trabalho são artigos científicos.

4.2.2 BLMSumm

O BLMSumm é um sumarizador automático proposto por Oliveira e Guelpeli (2011) que tem como característica a independência do idioma e do domínio. Para gerar os sumários o software permite fazer combinações entre diferentes métodos de classificação de sentenças. A Figura 18 apresenta o modelo do BLMSumm.

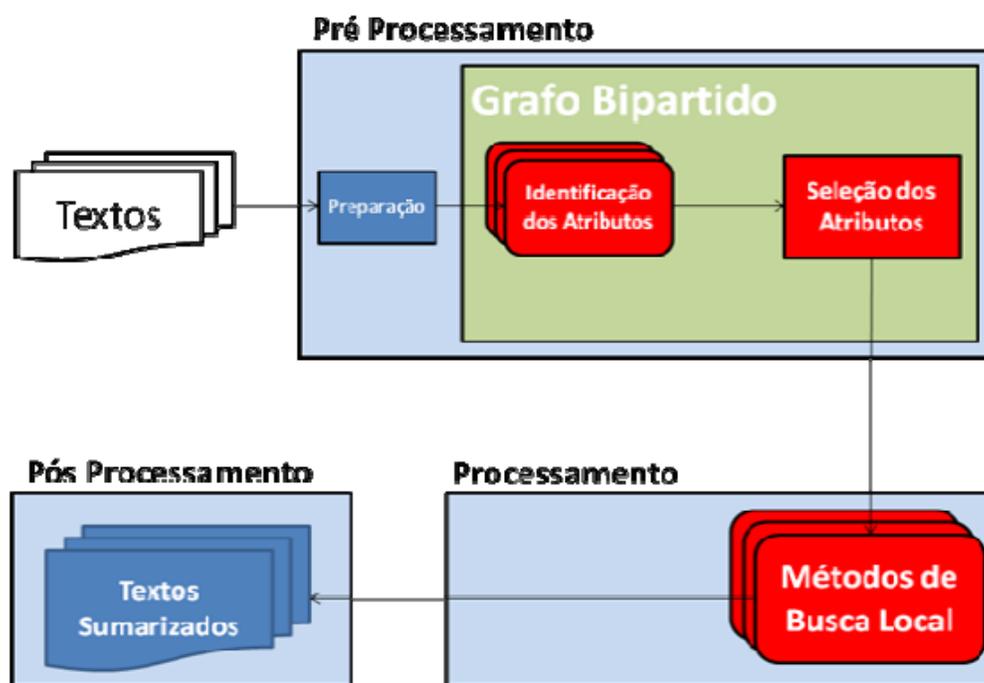


Figura 18. Modelo do BLMSumm (Oliveira e Guelpeli, 2011).

Neste trabalho todos os sumários gerados pelo BLMSumm foram feitos a partir da combinação do método de classificação de sentenças *TF-ISF* com o algoritmo Subida de encosta. A escolha destes métodos se deu por apresentarem o melhor comportamento na sumarização do corpus criado para teste, principalmente na taxa de compressão de 90%.

4.3 Processo de Sumarização Automática

O procedimento de SA pode ser exposto em duas etapas: a primeira na sumarização dos textos propriamente dita, e a segunda na avaliação dos sumários de forma automatizada com a ferramenta ROUGE.

A sumarização dos textos foi feita de acordo com quatro taxas de compressão diferentes: 90%, 80%, 70% e 50%, isto é, foram gerados sumários, respectivamente, com 10%, 20%, 30% e 50% de tamanho em relação ao texto original. No total, foram feitas 8 mil sumarizações (500 textos fonte * 4 taxas de compressão * 4 sumarizadores). A sumarização com o PragmaSUM foi feita de duas formas diferentes, a primeira utilizando as palavras chave de cada artigo para personificar o sumário, e a segunda sem a utilização destas palavras, ou seja, sem a utilização do perfil do usuário.

Na segunda etapa utilizou-se a ROUGE, que foi apresentado na seção 2.4 deste trabalho. A ferramenta gera planilhas individuais para as métricas *F-Measure*, *Precision* e *Recall* com o resultado das comparações individuais entre cada sumário automático com seu respectivo sumário manual e também planilhas com médias gerais, neste caso, médias de resultados dos 500 sumários avaliados para cada sumarizador. As comparações foram feitas entre cada um dos dez domínios de conhecimento.

5. RESULTADOS

Este capítulo apresenta os resultados das simulações feitas com o *Corpus* criado para os testes e os sumarizadores automáticos avaliados. Os resultados foram divididos por domínio e subdivididos pelas quatro taxas de compressão utilizadas na sumarização. Foram gerados tabelas e gráficos de *Recall*, *Precision* e *F-Measure*, mas devido ao grande volume de dados gerados pelas medidas usadas para aferir a eficiência dos sumários e o objetivo de se comparar a precisão que o PragmaSUM obtém com a utilização das palavras chave do texto, serão apresentados somente os gráficos de comparação entre as métricas *Precision* entre os resultados obtidos pelas quatro taxas de compressão.

A avaliação do ROUGE, compara sumários manuais com os sumários automáticos e gera as tabelas com os resultados do *Precision*, que compara o quanto do sumário manual coincide com o sumário automático, o que a torna a métrica mais recomendada para análise do efeito do uso das palavras chaves dos artigos na sumarização automática de textos com o PragmaSUM.

Os resultados variam numa escala que varia de 0 a 1 que quanto mais próximos de 1 estiverem, melhores serão os resultados obtidos e quanto mais se aproximarem do 0, piores serão os resultados. As tabelas e gráficos do *Precision*, *Recall* e *F-Measure* e respectivas comparações serão apresentadas no apêndice A. No fim de cada seção será mostrada uma comparação entre os resultados obtidos pelas quatro taxas de compressão aplicadas. Serão apresentadas, ainda, comparações entre os valores alcançados por cada sumarizador para todos os domínios, a comprovação da hipótese e discussão dos resultados.

5.1.1 Domínio Administração

A Tabela 3 e a Figura 19 mostram uma comparação feita entre as taxas de compressão das médias gerais do *Precision* alcançadas pelos sumarizadores automáticos para os textos do domínio Administração. O PragmaSUM obteve os melhores resultados, e ampliou a vantagem sobre sua versão sem chaves com valores maiores de compressão. O

GistSumm obteve os piores resultados, exceto quando a compressão foi de 90%, quando ultrapassou o BLMSumm.

Tabela 3. Comparação entre os resultados de *Precision* do domínio Administração.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,8888	0,79366	0,91603	0,9093
70%	0,82124	0,78366	0,88092	0,86515
80%	0,75515	0,7479	0,84415	0,8249
90%	0,63253	0,64673	0,76427	0,73987

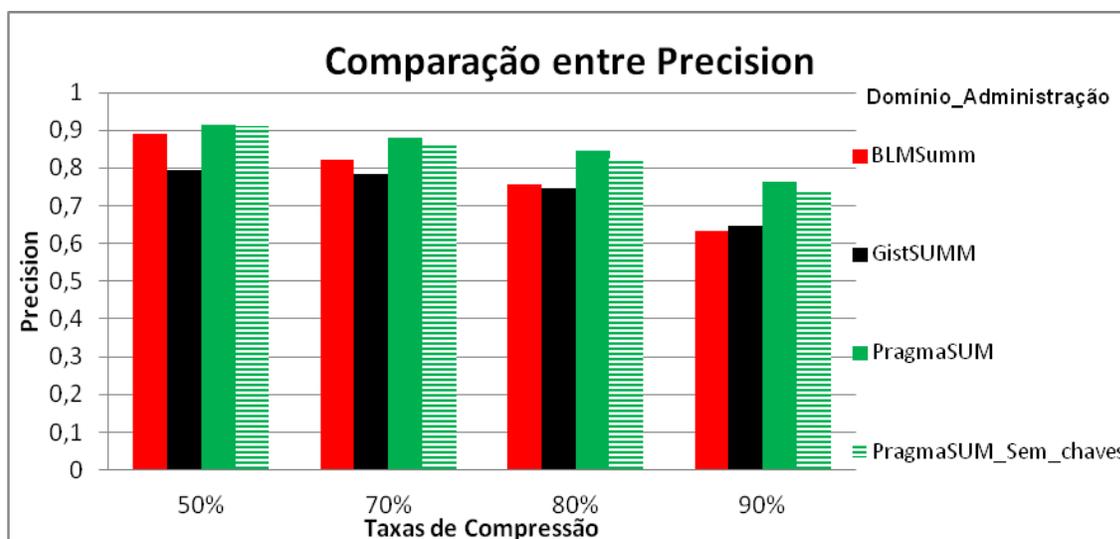


Figura 19. Comparação entre os resultados de *Precision* do domínio Administração.

5.1.2 Domínio Agronomia

A Tabela 4 e a Figura 20 mostram uma comparação feita entre as taxas de compressão das médias gerais do *Precision* alcançadas pelos sumarizadores automáticos para os textos do domínio Agronomia. O PragmaSUM obteve os melhores resultados quando as compressões foram 80% e 90%, com 50% e 70% a sua versão sem chaves teve leve vantagem. O GistSumm obteve os piores resultados, mas se aproximou do BLMSumm à medida em que a taxa compressão aumentou.

Tabela 4. Comparação entre os resultados de *Precision* do domínio Agronomia.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,8402	0,65554	0,8768	0,88035
70%	0,73472	0,64171	0,80941	0,81219
80%	0,63662	0,59751	0,74753	0,73551
90%	0,4695	0,45468	0,61782	0,57549

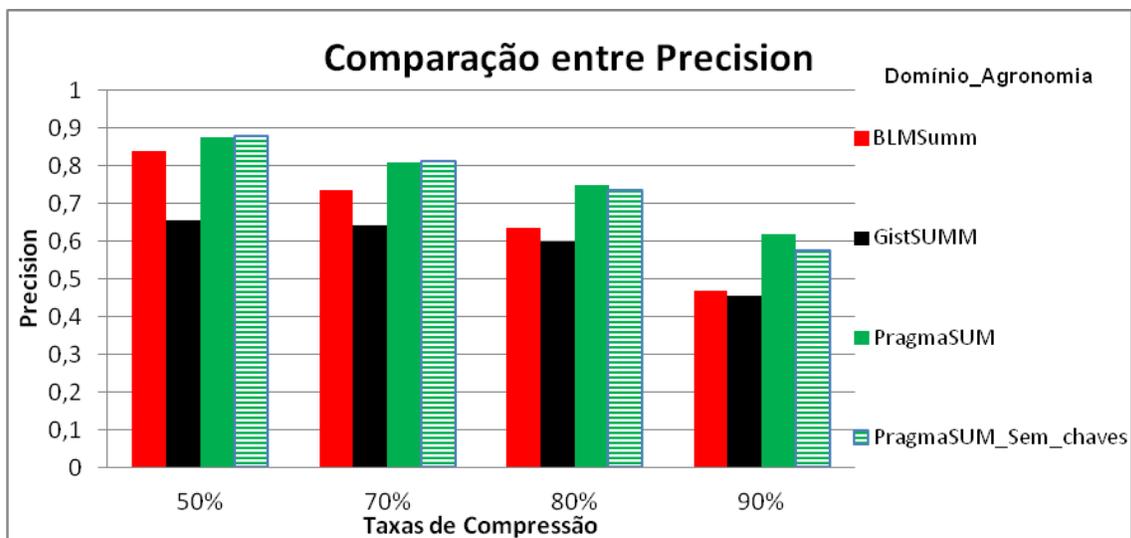


Figura 20. Comparação entre os resultados de *Precision* do domínio Agronomia.

5.1.3 Domínio Educação Física

A Tabela 5 e a Figura 21 mostram uma comparação feita entre as taxas de compressão das médias gerais do *Precision* alcançadas pelos sumarizadores automáticos para os textos do domínio Educação Física. O PragmaSUM obteve os melhores resultados com leve vantagem sobre a sua versão sem chaves. O GistSumm obteve os piores resultados, exceto quando a taxa de compressão foi 90%, quando ele ultrapassou o BLMSumm.

Tabela 5. Comparação entre os resultados de *Precision* do domínio Educação Física.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,86777	0,7467	0,89809	0,89319
70%	0,78361	0,7349	0,86464	0,85832
80%	0,71027	0,68868	0,82033	0,81386
90%	0,57054	0,57965	0,74635	0,71214

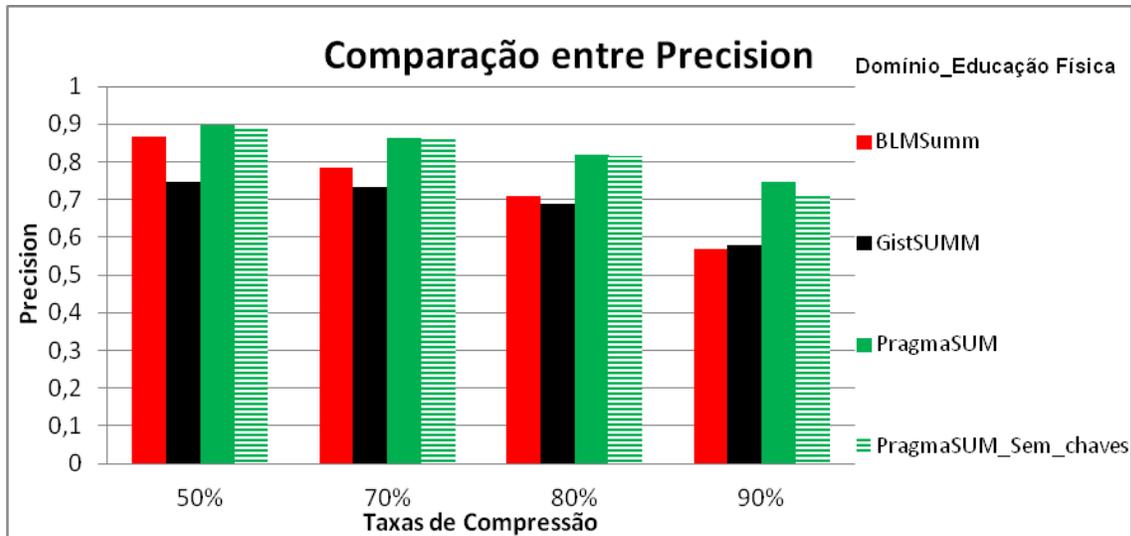


Figura 21. Comparação entre os resultados de *Precision* do domínio Educação Física.

5.1.4 Domínio Engenharia

A Tabela 6 e a Figura 22 mostram uma comparação feita entre as taxas de compressão das médias gerais do *Precision* alcançadas pelos sumarizadores automáticos para os textos do domínio Engenharia. A versão sem chaves do PragmaSUM obteve os melhores resultados, com ligeira vantagem sobre sua outra versão. O GistSumm obteve os piores resultados, exceto quando a taxa de compressão foi 90%, quando ele teve ligeira vantagem sobre o BLMSumm.

Tabela 6. Comparação entre os resultados de *Precision* do domínio Engenharia.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,8719	0,72004	0,90354	0,90865
70%	0,80894	0,7142	0,87489	0,87875
80%	0,73183	0,69611	0,84401	0,84718
90%	0,60009	0,60466	0,76382	0,77236

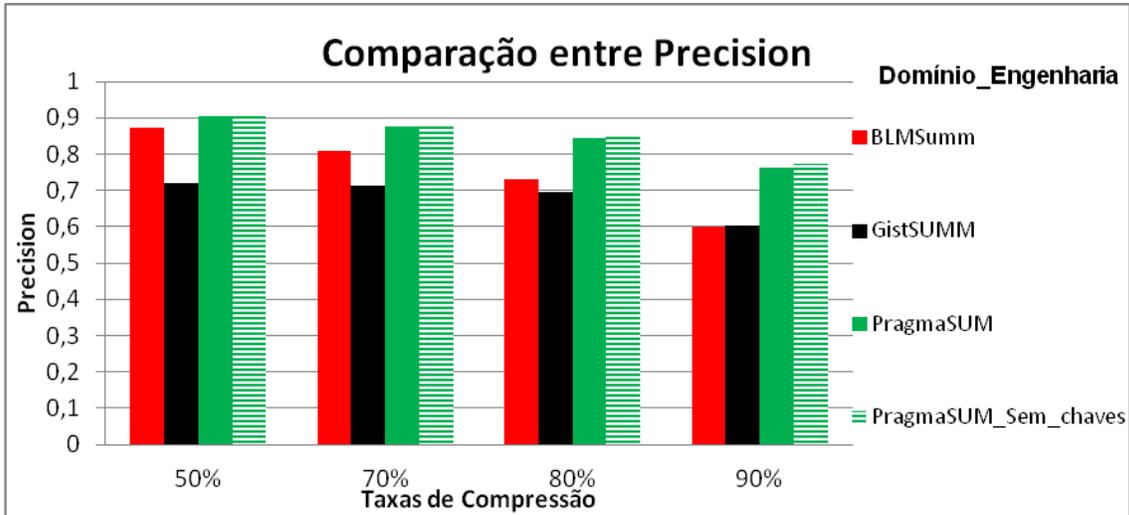


Figura 22. Comparação entre os resultados de *Precision* do domínio Engenharia.

5.1.5 Domínio Filosofia

A Tabela 7 e a Figura 23 mostram uma comparação feita entre as taxas de compressão das médias gerais do *Precision* alcançadas pelos sumarizadores automáticos para os textos do domínio Filosofia. O PragmaSUM obteve os melhores resultados, com boa vantagem. O GistSumm obteve os piores resultados.

Tabela 7. Comparação entre os resultados de *Precision* do domínio Filosofia.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,82195	0,74733	0,8522	0,83615
70%	0,7523	0,72787	0,817	0,79745
80%	0,68261	0,67582	0,78883	0,75485
90%	0,57852	0,57569	0,71941	0,68674

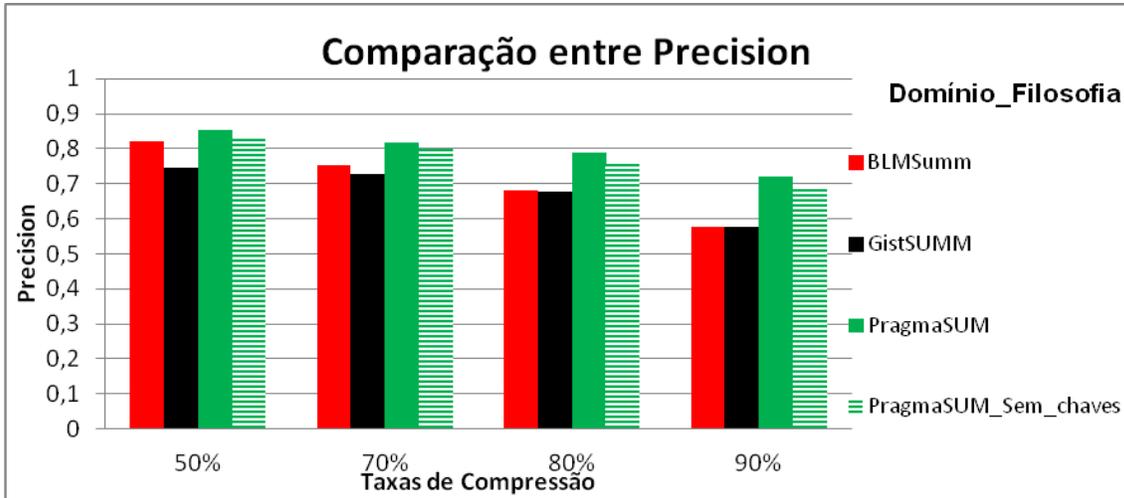


Figura 23. Comparação entre os resultados de *Precision* do domínio Filosofia.

5.1.6 Domínio Fisioterapia

A Tabela 8 e a Figura 24 mostram uma comparação feita entre as taxas de compressão das médias gerais do *Precision* alcançadas pelos sumarizadores automáticos para os textos do domínio Fisioterapia. O PragmaSUM obteve os melhores resultados, com boa vantagem. O GistSumm obteve os piores resultados, exceto quando a compressão foi 90%, quando ele teve leve vantagem sobre o BLMSumm.

Tabela 8. Comparação entre os resultados de *Precision* do domínio Fisioterapia.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,84055	0,65235	0,87466	0,86192
70%	0,71061	0,65051	0,79785	0,77221
80%	0,62178	0,61129	0,72907	0,68788
90%	0,44173	0,44883	0,5876	0,51953

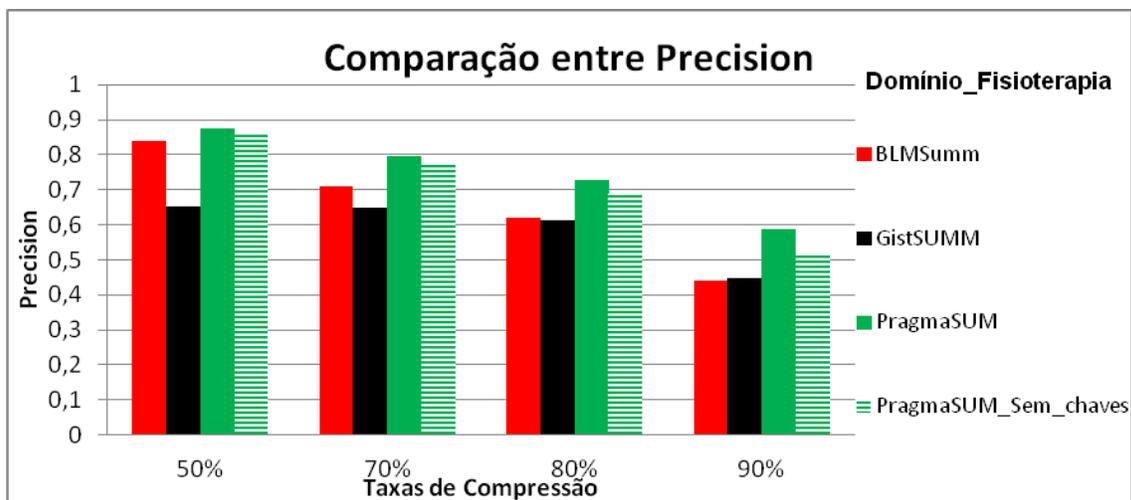


Figura 24. Comparação entre os resultados de *Precision* do domínio Fisioterapia.

5.1.7 Domínio Geografia

A Tabela 9 e a Figura 25 mostram uma comparação feita entre as taxas de compressão das médias gerais do *Precision* alcançadas pelos sumarizadores automáticos para os textos do domínio Geografia. O PragmaSUM obteve os melhores resultados, com boa vantagem. O GistSumm obteve os piores resultados, exceto quando a compressão foi de 80% e 90%, quando ele teve leve vantagem sobre o BLMSumm.

Tabela 9. Comparação entre os resultados de *Precision* do domínio Geografia.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,86642	0,75347	0,90368	0,89426
70%	0,77601	0,74222	0,86345	0,82431
80%	0,69616	0,69757	0,81836	0,77992
90%	0,55813	0,57906	0,72986	0,65509

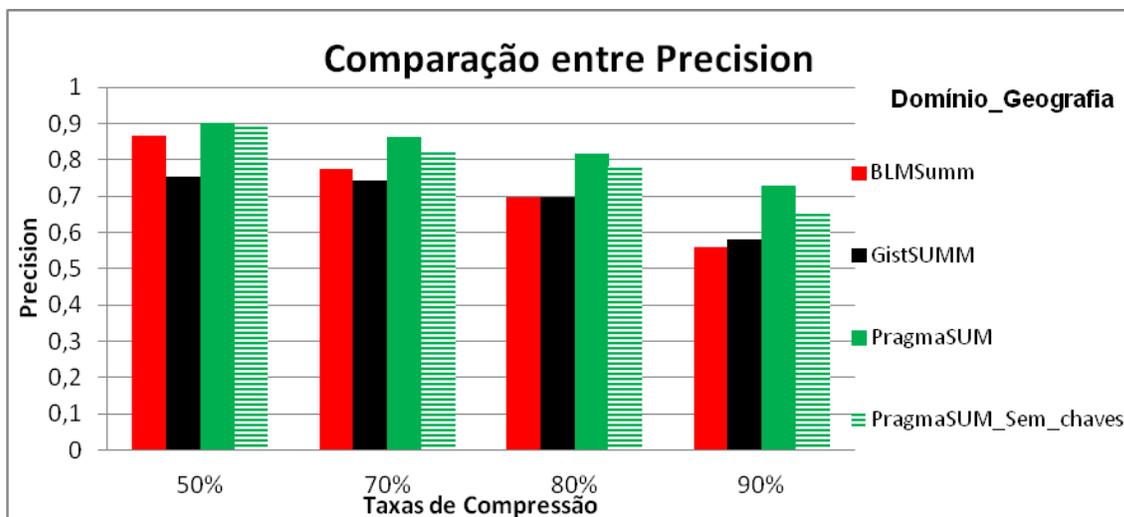


Figura 25. Comparação entre os resultados de *Precision* do domínio Geografia.

5.1.8 Domínio História

A Tabela 10 e a Figura 26 mostram uma comparação feita entre as taxas de compressão das médias gerais do *Precision* alcançadas pelos sumarizadores automáticos para os textos do domínio História. O PragmaSUM obteve os melhores resultados, com ligeira vantagem sobre sua versão sem chaves. O GistSumm obteve os piores resultados quando as compressões foram de 50% e 70%, e o BLMSumm os piores quando a compressão foi de 80% e 90%.

Tabela 10. Comparação entre os resultados de *Precision* do domínio História.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,845	0,77344	0,8774	0,87338
70%	0,75894	0,75827	0,83957	0,83483
80%	0,69714	0,72001	0,79955	0,79716
90%	0,58592	0,62808	0,73259	0,69771

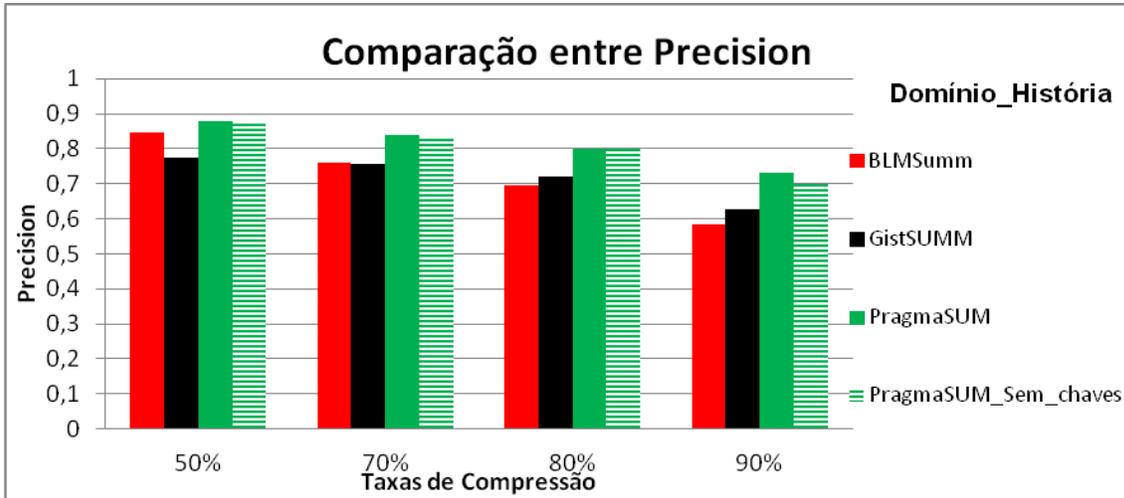


Figura 26. Comparação entre os resultados de *Precision* do domínio História.

5.1.9 Domínio Medicina

A Tabela 11 e a Figura 27 mostram uma comparação feita entre as taxas de compressão das médias gerais do *Precision* alcançadas pelos sumarizadores automáticos para os textos do domínio Medicina. O PragmaSUM obteve os melhores resultados, exceto quando a taxa de compressão foi 80%, quando sua versão sem chaves obteve os mesmos resultados. O GistSumm obteve os piores resultados exceto quando a compressão dos textos foi de 90%, onde obteve ligeira vantagem sobre o BLMSumm.

Tabela 11. Comparação entre os resultados de *Precision* do domínio Medicina.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,80604	0,70122	0,84976	0,84527
70%	0,70487	0,66985	0,7861	0,77244
80%	0,61346	0,60983	0,71372	0,71372
90%	0,44042	0,45428	0,58081	0,54125

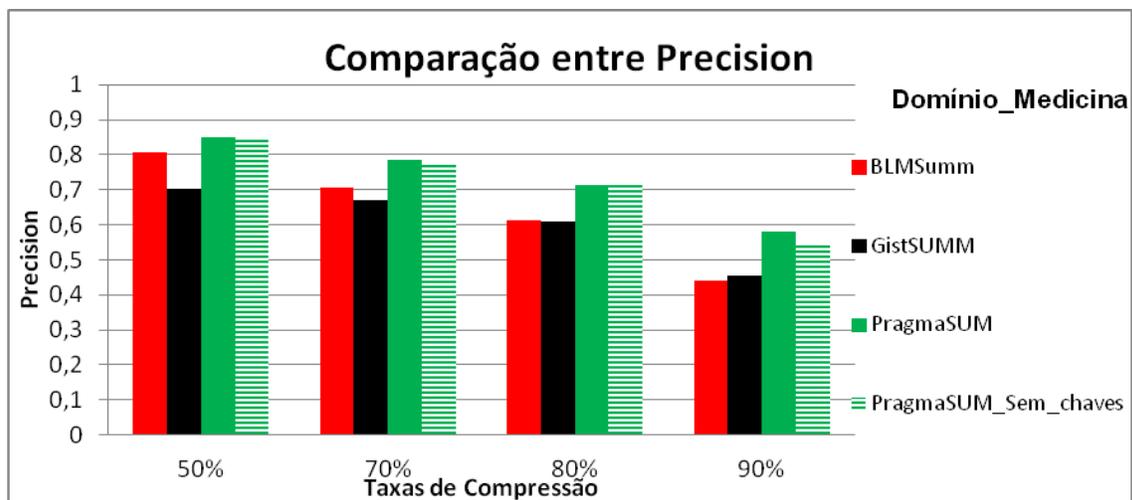


Figura 27. Comparação entre os resultados de *Precision* do domínio Medicina.

5.1.91 Domínio Psicologia

A Tabela 12 e a Figura 28 Apresentam uma comparação feita entre as taxas de compressão das médias gerais do *Precision* alcançadas pelos sumarizadores automáticos para os textos do domínio Psicologia. O PragmaSUM obteve os melhores resultados, com boa vantagem sobre os demais. O GistSumm obteve os piores resultados exceto quando a compressão dos textos foi de 80% e 90%, onde obteve ligeira vantagem sobre o BLMSumm.

Tabela 12. Comparação entre os resultados de *Precision* do domínio Psicologia.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,85757	0,76213	0,88959	0,88076
70%	0,76809	0,73778	0,85436	0,82675
80%	0,70087	0,70345	0,81329	0,76467
90%	0,57202	0,60603	0,74018	0,64036

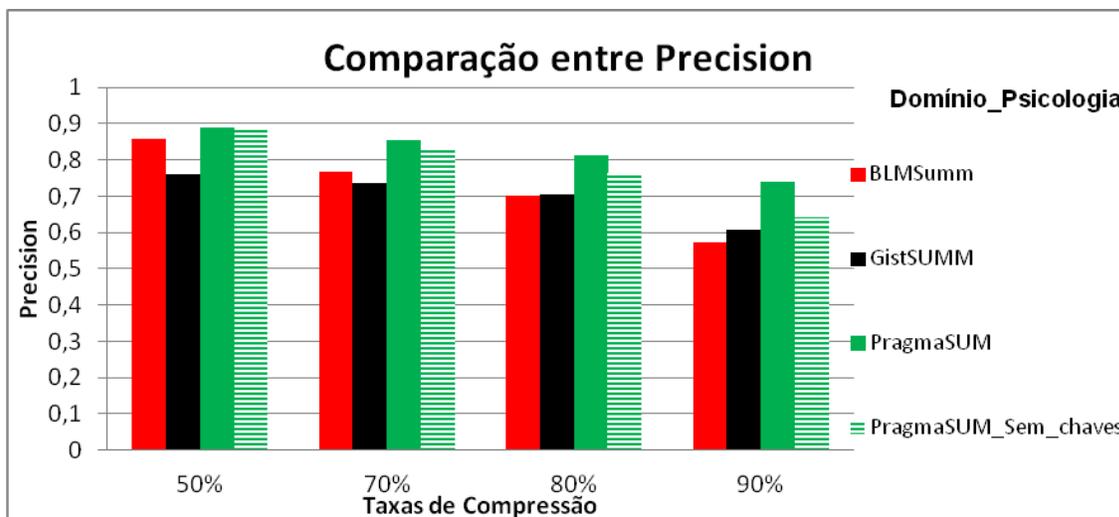


Figura 28. Comparação entre os resultados de *Precision* do domínio Psicologia.

5.2 Comparação entre os Domínios

5.2.1 Taxa de compressão 50%

Para taxa de compressão de 50% o domínio Administração obteve os melhores resultados do Precision para todos os sumarizadores automáticos avaliados, exceto para a versão sem chaves do PragmaSUM que alcançou o maior valor no domínio Engenharia, como pode se observar na Tabela 13 e na Figura 29.

Tabela 13. Comparação de *Precision* entre todos os domínios com taxa de compressão de 50%.

	BLMSumm	GistSumm	PragmaSUM	PragmaSUM_Sem_chaves
Administração	0,8888	0,79366	0,91603	0,9093
Agronomia	0,8402	0,65554	0,8768	0,88035
Educação Física	0,86777	0,7467	0,89809	0,89319
Engenharia	0,8719	0,72004	0,90354	0,90865
Filosofia	0,82195	0,74733	0,8522	0,83615
Fisioterapia	0,84055	0,65235	0,87466	0,86192
Geografia	0,86642	0,75347	0,90368	0,89426
História	0,845	0,77344	0,8774	0,87338
Medicina	0,80604	0,70122	0,84976	0,84527
Psicologia	0,85757	0,76213	0,88959	0,88076

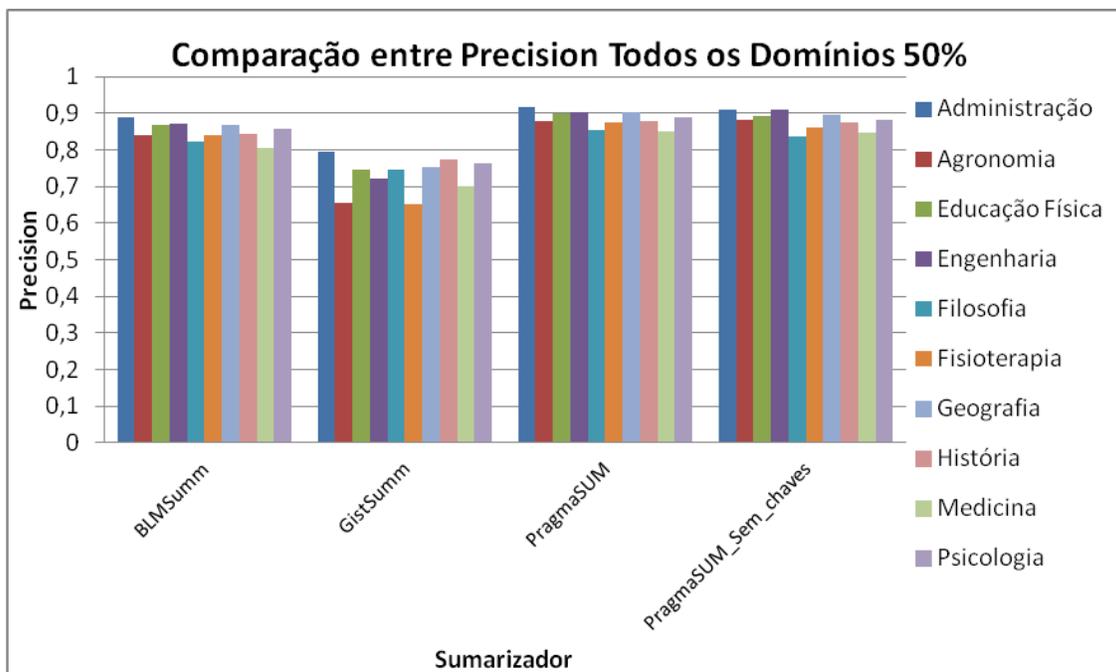


Figura 29. Comparação de *Precision* entre todos os domínios com taxa de compressão de 50%.

5.2.2 Taxa de compressão 70%

Para taxa de compressão de 70% o domínio Administração obteve os melhores resultados do Precision para todos os sumarizadores automáticos avaliados, exceto para a versão sem chaves do PragmaSUM que alcançou o maior valor no domínio Engenharia, repetindo o comportamento de quando a compressão foi de 50%, como pode-se observar na Tabela 14 e na Figura 30.

Tabela 14. Comparação de *Precision* entre todos os domínios com taxa de compressão de 70%.

	BLMSumm	GistSumm	PragmaSUM	PragmaSUM_Sem_chaves
Administração	0,82124	0,78366	0,88092	0,86515
Agronomia	0,73472	0,64171	0,80941	0,81219
Educação Física	0,78361	0,7349	0,86464	0,85832
Engenharia	0,80894	0,7142	0,87489	0,87875
Filosofia	0,7523	0,72787	0,817	0,79745
Fisioterapia	0,71061	0,65051	0,79785	0,77221
Geografia	0,77601	0,74222	0,86345	0,82431
História	0,75894	0,75827	0,83957	0,83483
Medicina	0,70487	0,66985	0,7861	0,77244
Psicologia	0,76809	0,73778	0,85436	0,82675

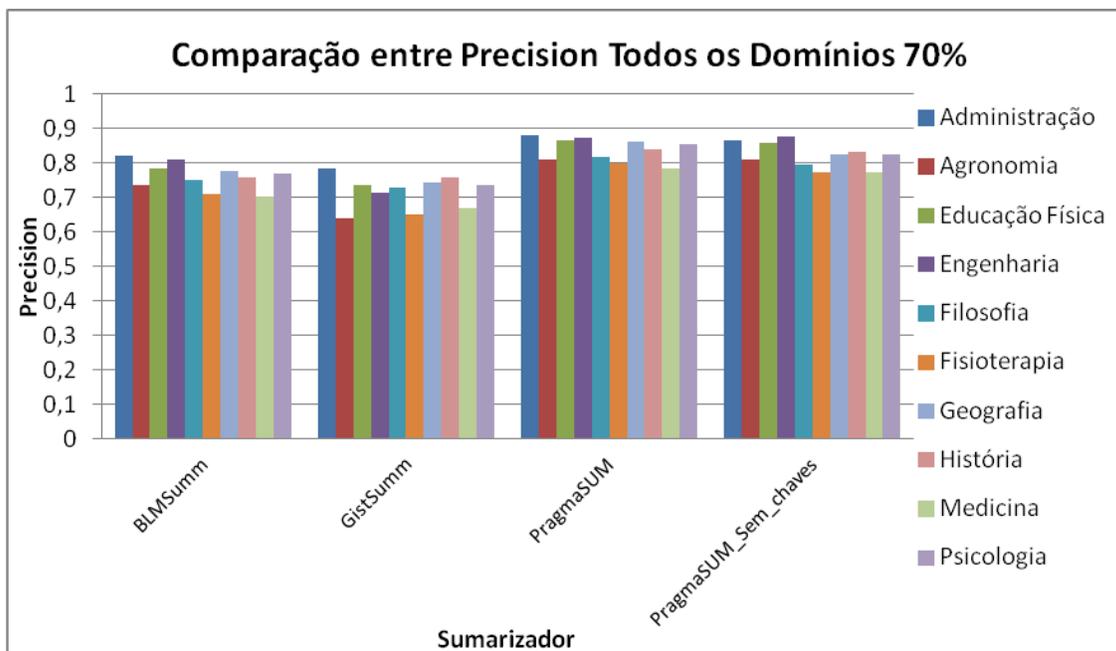


Figura 30. Comparação de *Precision* entre todos os domínios com taxa de compressão de 70%.

5.2.3 Taxa de compressão 80%

Para taxa de compressão de 80% o domínio Engenharia obteve o melhor resultado do *Precision* para a versão sem chaves do PragmaSUM, o domínio de Administração conseguiu os melhores resultados dos demais sumarizadores, como pode-se observar na Tabela 15 e na Figura 31.

Tabela 15. Comparação de *Precision* entre todos os domínios com taxa de compressão de 80%.

	BLMSumm	GistSumm	PragmaSUM	PragmaSUM_Sem_chaves
Administração	0,75515	0,7479	0,84415	0,8249
Agronomia	0,63662	0,59751	0,74753	0,73551
Educação Física	0,71027	0,68868	0,82033	0,81386
Engenharia	0,73183	0,69611	0,84401	0,84718
Filosofia	0,68261	0,67582	0,78883	0,75485
Fisioterapia	0,62178	0,61129	0,72907	0,68788

Geografia	0,69616	0,69757	0,81836	0,77992
História	0,69714	0,72001	0,79955	0,79716
Medicina	0,61346	0,60983	0,71372	0,71372
Psicologia	0,70087	0,70345	0,81329	0,76467

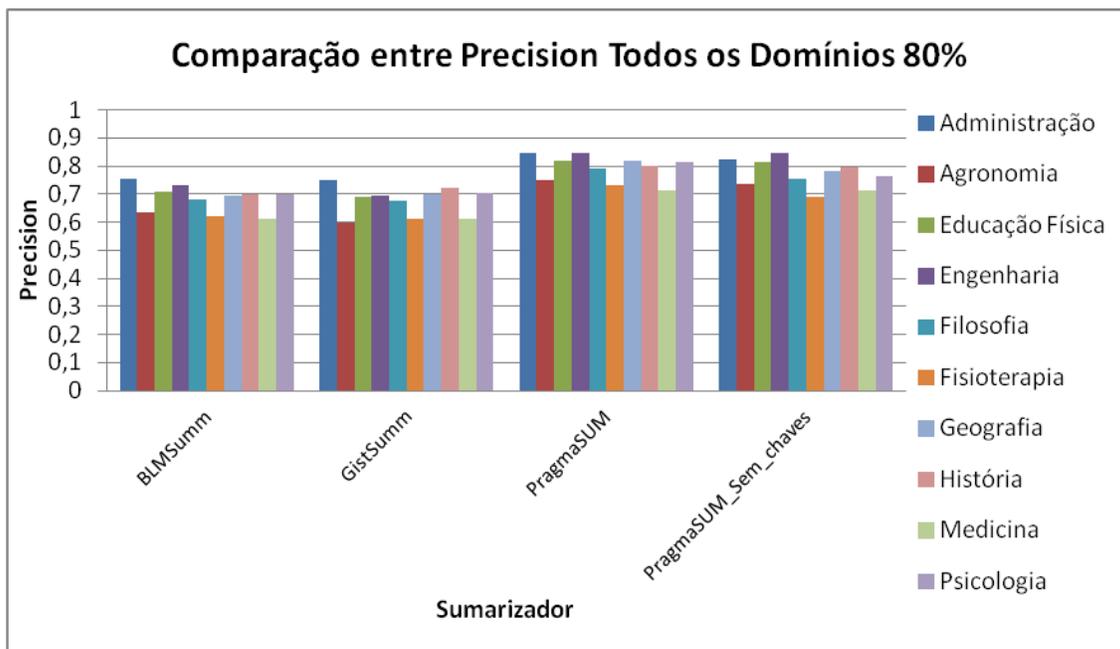


Figura 31. Comparação de *Precision* entre todos os domínios com taxa de compressão de 80%.

5.2.4 Taxa de compressão 90%

Para taxa de compressão de 90% o domínio Engenharia obteve o melhor resultado do *Precision* para a versão sem chaves do PragmaSUM, o domínio de Administração demonstrou os melhores resultados dos demais sumarizadores, como pode-se observar na Tabela 16 e na Figura 32.

Tabela 16. Comparação de *Precision* entre todos os domínios com taxa de compressão de 90%.

	BLMSumm	GistSumm	PragmaSUM	PragmaSUM_Sem_chaves
Administração	0,63253	0,64673	0,76427	0,73987
Agronomia	0,4695	0,45468	0,61782	0,57549
Educação Física	0,57054	0,57965	0,74635	0,71214
Engenharia	0,60009	0,60466	0,76382	0,77236
Filosofia	0,57852	0,57569	0,71941	0,68674
Fisioterapia	0,44173	0,44883	0,5876	0,51953
Geografia	0,55813	0,57906	0,72986	0,65509
História	0,58592	0,62808	0,73259	0,69771
Medicina	0,44042	0,45428	0,58081	0,54125
Psicologia	0,57202	0,60603	0,74018	0,64036

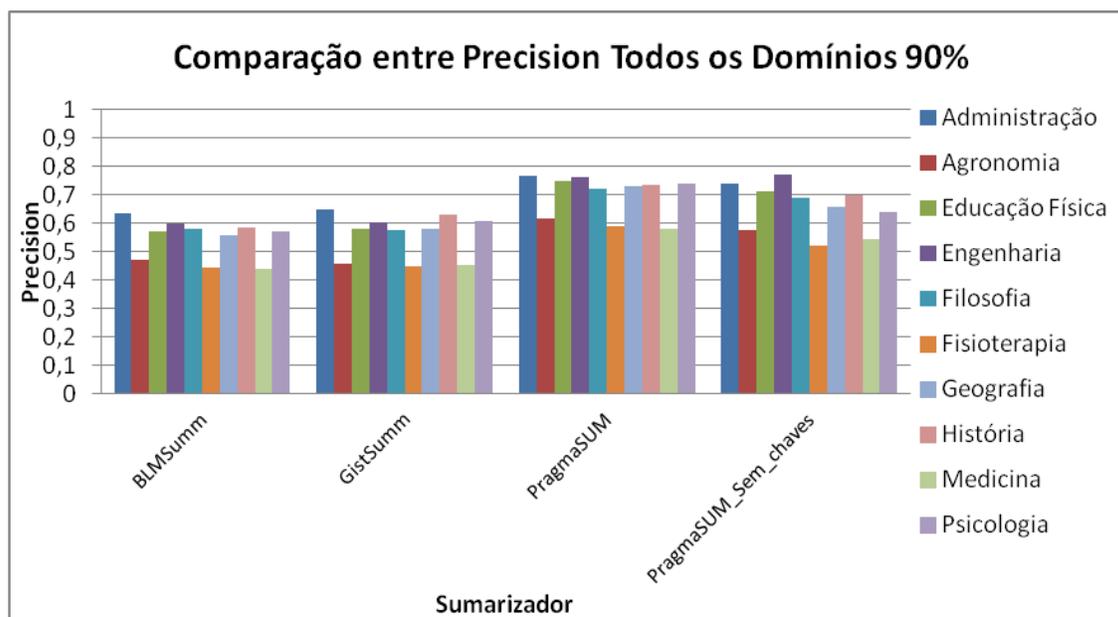


Figura 32. Comparação de *Precision* entre todos os domínios com taxa de compressão de 90%.

5.3 Análise dos testes estatísticos

Diante dos resultados apresentados foram utilizados os testes estatísticos ANOVA de Friedman e Coeficiente de concordância de Kendall, descritos na seção 2.5, para verificar se existe uma diferença significativa entre as amostras. Estes resultados estatísticos também foram divididos por domínio e por taxa de compressão e, devido ao grande volume de dados, serão apresentados no apêndice B. Para validação destes resultados com os testes estatísticos foi utilizado uma versão trial do software StatPlus®, disponível em <http://www.analystsoft.com/en/products/statplus/>.

Os resultados dos testes estatísticos Coeficiente de concordância de Kendall e Ordem médio demonstrados em todas as tabelas do apêndice C foram satisfatórios, o que permite afirmar que houve diferença significativa entre as amostras experimentadas, ou seja, todos os resultados apresentados anteriormente foram confirmados.

Pode-se afirmar que o PragmaSUM, com a utilização das palavras chave, obteve sucesso em 33 das 40 amostras analisadas, ou seja, 82.5% de sucesso. O PragmaSUM foi superado por sua versão sem chaves em todas as taxas de compressão do domínio engenharia, e quando as taxas de compressão foram 50% e 70% no domínio agronomia, além de ter valores idênticos quando a taxa de compressão foi de 80% no domínio medicina.

5.4 Análise-Hipótese

A hipótese deste trabalho consiste na possibilidade da criação de um sumário automático que consiga sumarizar um texto fonte independente de seu idioma e de seu domínio de conhecimento, que se adapte ao perfil do leitor para gerar sumários mais personificados e com o tamanho desejado pelo usuário, mantendo ou melhorando o desempenho que os sumários existentes conseguem alcançar.

A hipótese nula deste trabalho consiste na afirmação que a utilização de palavras chave na sumarização de textos não pode gerar um sumário com mais precisão e personificado de acordo com o perfil do leitor.

$$H_0 = K_{\text{Sumarização_com_palavras_chave}} < K_{\text{Sumarização}} \quad (12)$$

Onde:

H_0 = hipótese nula;

$K_{\text{Sumarização_com_palavras_chave}}$ = K amostras de sumários gerados com o PragmaSUM com utilização de palavras chave na sumarização;

$K_{\text{sumarização}}$ = K amostras de sumários automáticos que não utilizam palavras chave na sua criação;

Quando a hipótese nula for rejeitada, outra hipótese, a alternativa H_1 deve ser aceita, ou seja, a performance do sumarizador automático aqui apresentado seja maior do que os sumarizadores já existentes na literatura.

$$H_1 = K_{\text{Sumarização_com_palavras_chave}} > K_{\text{Sumarização}} \quad (13)$$

As hipóteses foram baseadas nas amostras obtidas na ferramenta ROUGE dos textos fonte do *corpus* criado para testes e analisados com diferentes graus de compressão e nos diferentes sumarizadores automáticos, com as medidas de *Recall*, *Precision* e *F-Measure*.

Para comprovação foi utilizado o teste ANOVA de Friedman, que considera que as diversas amostras são, estatisticamente, idênticas na sua distribuição (hipótese de nulidade, ou de H_0). A hipótese alternativa (H_1) aponta como elas são significativamente diferentes, na sua distribuição e o teste de concordância de Kendall normaliza o teste estatístico de Friedman, com a finalidade de gerar uma avaliação de concordância, ou não, com ranques estabelecidos (GUELPELI, 2012).

5.5 Discussão

Esta seção tem como objetivo discutir os resultados como um todo, pela análise das métricas *Recall*, *Precision* e *F-Measure*.

Na medida *Precision*, observou-se que os resultados obtidos pelas duas versões da sumarização com o PragmaSUM foram os mais satisfatórios. O método de sumarização com as palavras chave só foi superado no domínio Engenharia, mas ainda assim foi superior aos outros dois sumarizadores analisados. Notou-se ainda que com o aumento da taxa de compressão utilizada, a diferença entre os valores obtidos pelas duas versões do PragmaSUM e dos outros sumarizadores analisados foi maior.

Mediante os resultados obtidos com o PragmaSUM, a utilização das palavras chave na sumarização de textos melhorou o desempenho obtido na métrica *Precision*. Essa melhora foi observada principalmente com as maiores taxas de compressão aplicadas. Vale ressaltar que não houve nenhum tipo de análise prévia dos textos em relação a influência das palavras chave em seu conteúdo, o que pode influenciar nos resultados obtidos pelas métricas analisadas.

Um fator que merece atenção é a comparação entre os dois métodos utilizados na sumarização pelo PragmaSUM. Houve grande ganho na utilização das palavras chave na maioria dos domínios, sobretudo quando a taxa de compressão usada era 80% ou 90%. Com a diminuição do tamanho dos sumários a seleção de poucas sentenças se torna necessário, aumentando as chances de sentenças que contenham as palavras chave tomem o lugar de outras sentenças no sumário.

Analisando os valores de todas as tabelas do Apêndice C, os testes estatísticos Coeficiente de concordância de Kendall e Ordem médio são muito significativos, ou seja, os resultados apresentados na seção 5.1 foram confirmados. Pode-se concluir que a hipótese H_0 foi rejeitada e a hipótese H_1 foi aceita em 82.5% dos casos.

6. CONCLUSÃO

O principal objetivo desse trabalho foi apresentar o sumarizador automático de textos, PragmaSUM. Para isso foi criado um *corpus* para testes formado por artigos científicos na língua portuguesa e com dez domínios diferentes foram avaliados o BLMSumm, o GistSumm, e duas formas de sumarização do PragmaSUM, com e sem a utilização das palavras chave.

Nos testes realizados foram gerados sumários com quatro graus de compressão diferentes, com 50%,70%,80% e 90%. Os resultados foram avaliados pela ferramenta ROUGE com a métricas *Recall*, *Precision* e *F-Measure*.

Uma comparação feita entre os resultados dos domínios, mostrou que de acordo com a métrica *Precision* o domínio Administração obteve os melhores resultados, com todos os sumarizadores, com exceção da versão sem chaves do PragmaSUM que apresentou o domínio Engenharia com os resultados mais altos. O PragmaSUM só foi superado no domínio engenharia e quando a taxa de compressão foi de 50% e 70% no domínio agronomia, onde sua versão sem chaves teve ligeira vantagem.

Os resultados apresentados foram confirmados com o uso dos testes estatísticos ANOVA de Friedman e Coeficiente de concordância de Kendall.

Observando as quatro taxas de compressão aplicadas percebe-se que os sumarizadores automáticos obtiveram, em ambos os domínios, resultados mais homogêneos quando a compressão foi de 50% e muita variação quando a taxa foi de 90%. Observou-se, também, conforme demonstrado nos resultados que os sumarizadores automáticos apresentaram uma tendência: quanto maior a taxa de compressão aplicada, pior eram os resultados, com grande variação, com exceção do PragmaSUM, que não teve grande perda com o aumento da compressão.

Como dito anteriormente, não houve nenhum tipo de análise da influência das palavras chave no conteúdo do texto. Como o domínio administração obteve os melhores resultados do *Precision*, pode significar que neste domínio as palavras chave têm maior importância no texto e nos domínios medicina e fisioterapia têm menor importância, já que eles obtiveram os resultados mais baixos.

Outro fator que pode ser considerado no desempenho de cada domínio é o seu tamanho. De maneira geral, os domínios maiores obtiveram os melhores resultados, e os menores, os piores. Isto pode ser relacionado com o fato de que com maiores taxas de compressão há maior perda nos resultados, já que textos menores, conseqüentemente, geram sumários menores.

Como dito anteriormente, o fato de o PragmaSUM se sair bem com as taxas de compressão mais elevadas com o uso de palavras chave, é justamente por que seu sumário é formado por um maior número de sentenças importantes do texto, ou seja, sentenças que contém as palavras chave disponibilizadas pelo autor.

A partir dessas constatações, nas próximas seções serão apresentadas as contribuições deste trabalho, assim como suas limitações e possíveis trabalhos futuros.

6.1 Contribuições

A partir dos experimentos e das análises dos resultados dos sumários gerados pelo PragmaSUM, apresentados neste trabalho, podem-se destacar algumas contribuições importantes para a área de sumarização automática de textos:

- Criou-se um modelo que melhora a precisão do sumário.
- Criou-se sumários mais personificados, de acordo com preferências do usuário;
- Viabilizou-se um estudo sobre a importância das escolhas das palavras chaves em artigos científicos;
- Possibilitou-se uma estrutura algorítmica para uso do aprendizado de máquina;
- Viabilizou-se um estudo da importância da palavra chave para o processo de indexação de artigos científicos.

6.2 Limitações

Neste trabalho foram usados apenas textos em português; o *corpus* criado contém palavras chaves e resumos retirados de artigos científicos na língua portuguesa; acredita-se que existe a necessidade de testar em idiomas diferentes. Outro fator limitante são os domínios: apesar de ter dez domínios seria interessante aumentar estes domínios. Uma

terceira observação é que tipicamente usa-se no máximo cinco palavras chaves em artigo científico, caberia um estudo para verificar a eficiência ou não deste valor.

6.3 Trabalhos Futuros

Utilizar um corpus em diferentes idiomas e em mais domínios para uma avaliação mais ampla. Aprimorar o método de personificação do sumário pelas palavras chave usando aprendizado de máquina, e com isso eliminar interação humana no processo e possibilitar ao PragmaSUM o aprendizado do perfil linguístico do usuário.

Realizar um estudo sobre a relevância do uso das palavras chaves no conteúdo de artigos científicos, tornando possível uma análise de qual critério os pesquisadores utilizam para escolher as palavras chave em seus trabalhos, se elas realmente tem relevância no texto ou são escolhidas para direcionarem pesquisas realizadas por algoritmos de busca na internet para seus artigos, assim como um estudo da quantidade dessas palavras chaves na questão da indexação de textos científicos.

REFERENCIAS

ALUÍSIO, S.M.; ALMEIDA, G.M.B. O que é e como se constrói um corpus? Lições aprendidas na compilação de vários corpora para pesquisa lingüística. **Calidoscópico**, (UNISINOS). vol. 4, n. 3, p. 155-177, set/dez 2006.

CALLEGARI-JACQUES, S. M. **Bioestatística: Princípios e Aplicações**. Porto Alegre: Artmed, 2007.

DELGADO, C.H; VIANA C.E; GUELPELI, M.C.V; **Comparando sumários de referência humanos com extratos ideais no processo de avaliação de sumários extrativos**. In: IADIS Ibero-Americana WWW/Internet 2010, p. 293 – 300 Algarve, Portugal, 2010.

ESPINA, A.P.; RINO, L.H.M.; **Utilização de Métodos Extrativos na Sumarização Automática de Textos**; Série de Relatórios do NILC. NILC-TR-02-06, São Carlos, Março, 2002;

FERNANDES H. M.; GUELPELI, M. V. C. **Creación de corpus en lengua española para su utilización en testes acerca de Sumarización Automática**. In: 6th International Conference on Corpus Linguistics-CILC 2014, 2014, Las Palmas de Gran Canaria. 6th International Conference on Corpus Linguistics-CILC 2014, 2014.

Finecount (2.6.) [Programa de Computador]. Tilti Systems. Disponível em: <<http://www.tilti.com/software-for-translators/finecount/http://pt.shvoong.com/internet-and-technologies/software/1862866-finecount-contar-palavras-http-www/>>. Acesso em 12 mai. 2014.

GUELPELI, M.V.C; GARCIA, A.C.B; **Sumarizador Automático Baseado em Perfis Pragmáticos**; In: IADIS International Conference WWW/Internet 2007, Vila Real, ISBN: 978-972-8924-44-7, IADIS Press, v.II. p.149 – 153, Portugal, 2007.

GUELPELI, M.V.C; BERNARDINI, F.C; GARCIA, A.C.B; **Todas as Palavras da Sentença como Métrica para um sumarizador Automático**. In: Tecnologia da Informação e da Linguagem Humana-TIL, WebMedia, 2008. P. 287-291, Vila Velha, Brasil, 2008.

GUELPELI, M.V.C; **Cassiopeia: Um modelo de agrupamento de textos baseado em sumarização**. 2012. 220f. Tese (Doutorado em Computação) – Universidade Federal Fluminense, Rio de Janeiro, 2012.

LEITE, D.S.; RINO, L.H.M.; **Uma comparação entre sistemas de sumarização automática extrativa**; Departamento de Computação, UFSCar CP 676, 13565-905, São Carlos, Outubro, 2006;

LIN, C-Y. and HOVY, E.H. Automatic Evaluation of Summaries Using N-gram Co occurrence Statistics. **In the Proceedings of Language Technology Conference – HLT**. Edmonton, Canadá, 2003.

LUCHI, D.; RIBEIRO, E. **Sumarização Automática de Textos via Ranqueamento de Sentenças**. 2011. Universidade Federal do Espírito Santo – UFES.

LUHN, H. P. The Automatic Creation of Literature Abstracts. **IBM Journal of Research and Development**, vol. 2, 157-165, 1958.

MARTINS, C.B.; PARDO, T.A.S.; ESPINA, A.P.; RINO, L.H.M. (2001). **Introdução à Sumarização Automática**. Relatório Técnico RT-DC 002/2001. Departamento de Computação, Universidade Federal de São Carlos. São Carlos-SP, Fevereiro, 38p.

NOGUEIRA, B.M.; **Avaliação de métodos não-supervisionados de seleção de atributos para Mineração de Textos**. 2009. 104f. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) - Universidade de São Paulo, São Carlos, 2009.

OLIVEIRA, M. A; GUELPELI, M. V. C. **BLMSumm - Métodos de Busca Local e Metaheurísticas na Sumarização de Textos** In: Proceedings of the ENIA - VIII Encontro Nacional de Inteligência Artificial 2011, p. 287 - 298 Natal, Brasil, 2011.

OLIVEIRA, R.R.; GUELPELI, M.V.C. Building a Corpus in Italian Written Language. In: **6th International Conference on Corpus Linguistics (CILC2014)**. Las Palmas de Gran Canaria, Espanha, 2014. No prelo.

OLIVEIRA, R.R.; GUELPELI, M.V.C. Corpus in Italian of the Journalism and Medical Fields. In: **Second Asia Pacific Corpus Linguistics Conference (APCLC 2014)**, Abstract. Hong Kong, China, 2014 b. No prelo.

PARDO, T.A.S e RINO, L.H.M. (2002). **DMSumm: Um Gerador Automático de Sumários**. In Anais do I Workshop de Teses e Dissertações em Inteligência Artificial – WTDIA, pp. 1-10. Recife-PE, Brasil. 11 a 14 de Novembro.

PARDO, T.A.S; **GistSumm: Um Sumarizador Automático Baseado na Ideia Principal de Textos**; Série de Relatórios do NILC. NILC-TR-02-13, São Carlos, Setembro, 2002;

PARDO, T.A.S; **GistSumm – GIST SUMMarizer: Extensões e Novas Funcionalidades**; Série de Relatórios do NILC. NILC-TR-05-05, São Carlos, Fevereiro, 2005;

PARDO, T.A.S; **Sumarização Automática: Principais Conceitos e Sistemas para o Português Brasileiro**; Série de Relatórios do NILC. NILC-TR-08-04, São Carlos, Maio, 2008;

PARDO, T.A.S; RINO, L.H.M; **TeMário: Um Corpus para Sumarização Automática de Textos**; Série de Relatórios do NILC. NILC-TR-03-09, São Carlos, Outubro, 2003;

PARDO, T.A.S; RINO, L.H.M; **Descrição do GEI – Gerador de Extratos Ideais para o Português do Brasil**; Série de Relatórios do NILC. NILC-TR-04-07, São Carlos, Agosto, 2004;

PARDO, T.A.S; **SEnTER: Um Segmentador Sentencial Automático para o Português do Brasil**; Série de Relatórios do NILC. NILC-TR-06-01, São Carlos, Janeiro, 2006;

SARDINHA, T. B. **Linguística de Corpus**. Baruerí, SP: Manole, 2004.

StatPlus® (Versão trial) [Programa de Computador]. AnalystSoft Inc. Disponível em: <<http://www.analystsoft.com/en/products/statplus/>>. Acesso em jun. 2014.

APÊNDICES

APÊNDICE A – TABELAS E GRÁFICOS COM OS RESULTADOS DO F-MEASURE, PRECISION E RECALL

O apêndice A traz os resultados das Sumarizações Automáticas (SA) realizadas num corpus de textos em português com textos de 10 domínios de conhecimento.

1. Domínio Administração

Esta seção descreve os resultados das SA realizadas no domínio Administração.

1.1. Taxa de Compressão de 50%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio Administração. A taxa de compressão aplicada foi de 50%.

A figura a seguir (média acumulada de F-Measure) apresenta os resultados das SA para cada um dos sumarizadores automáticos. Observa-se que há bastante equilíbrio nos resultados, mas o GistSumm possui uma ligeira vantagem.

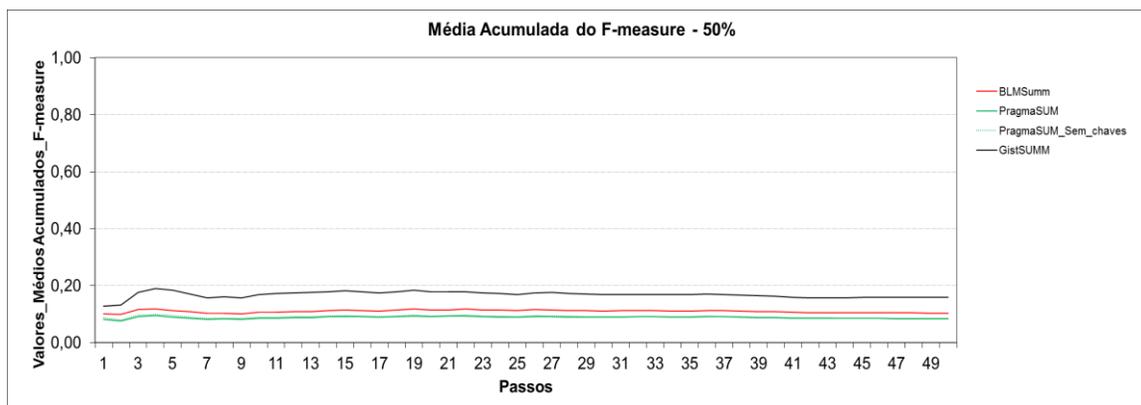


Figura 33. Média acumulada de F-Measure com taxa de compressão de 50% do domínio Administração.

Na Figura 34 (média acumulada de Precision) têm-se os resultados para cada um dos sumarizadores automáticos. Quanto à precisão (Precision) as duas versões do PragmaSUM apresentou os melhores resultados, sem muita diferença entre as duas versões.

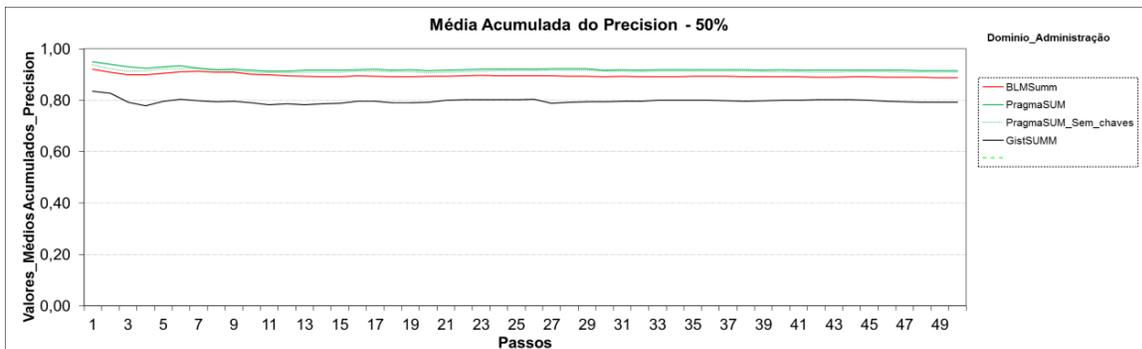


Figura 34. Média acumulada de Precision com taxa de compressão de 50% do domínio Administração.

A Figura 35 (média acumulada de Recall) mostram os resultados da cobertura para cada um dos sumarizadores automáticos. É possível perceber que existe uma variação muita pequena dentre resultados aferidos e o GistSumm possui ligeira vantagem.

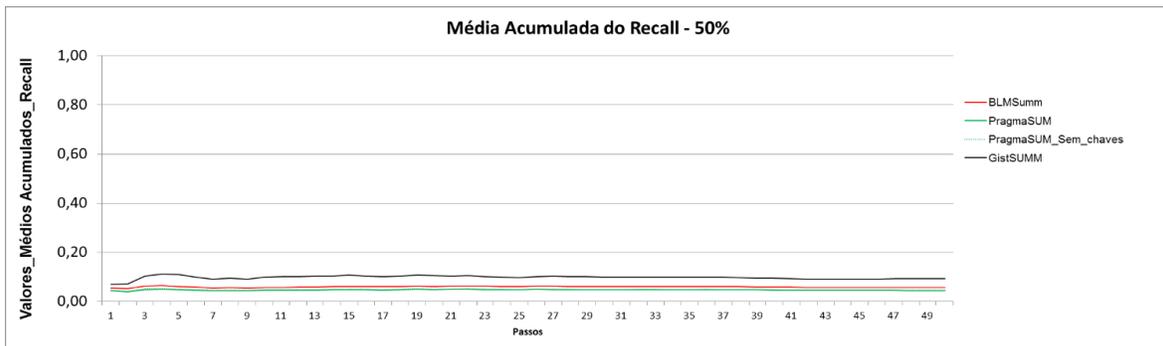


Figura 35. Média acumulada de Recall com taxa de compressão de 50% do domínio Administração.

1.2. Taxa de Compressão de 70%

É possível perceber na Figura 36 (média acumulada de F-Measure) que o GistSumm apresenta os resultados mais satisfatórios com uma mínima diferença para o BLMSumm e PragmaSUM.

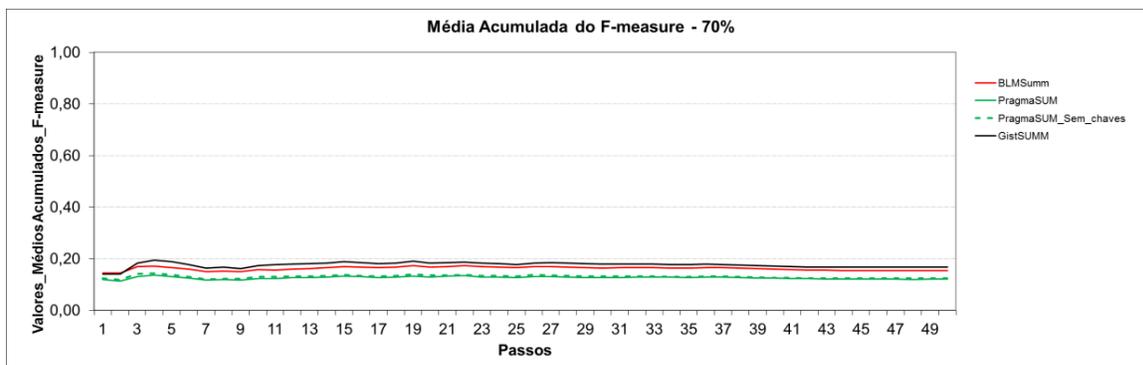


Figura 36. Média acumulada de F-Measure com taxa de compressão de 70% do domínio Administração.

Percebe-se na Figura 37 (média acumulada de Precision) que o sumarizador com os valores de precisão mais altos foi o PragmaSUM, seguido pela sua versão sem chaves e o com valores mais baixos foi o GistSumm.

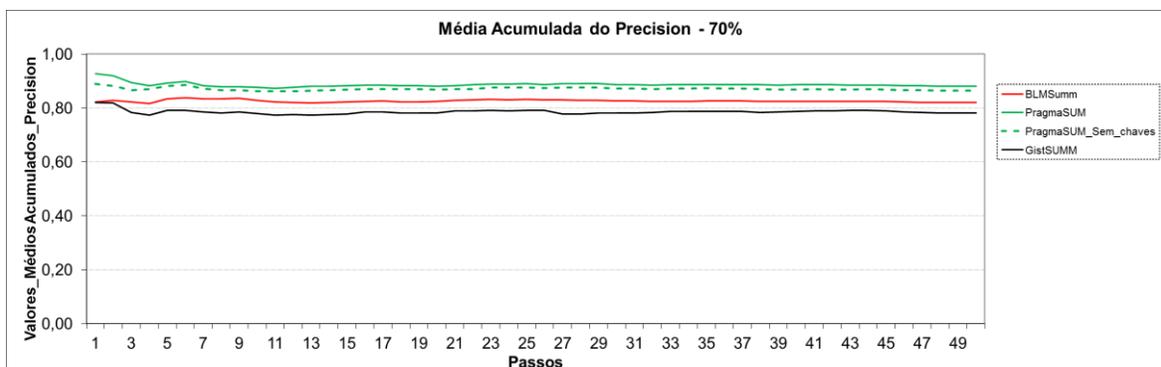


Figura 37. Média acumulada de Precision com taxa de compressão de 70% do domínio Administração.

A Figura 38 (média acumulada de Recall) mostram que quanto à cobertura os sumarizadores que obtiveram os melhores resultados foram o BLMSum e o GistSumm, mas com valores muito próximo do PragmaSUM.

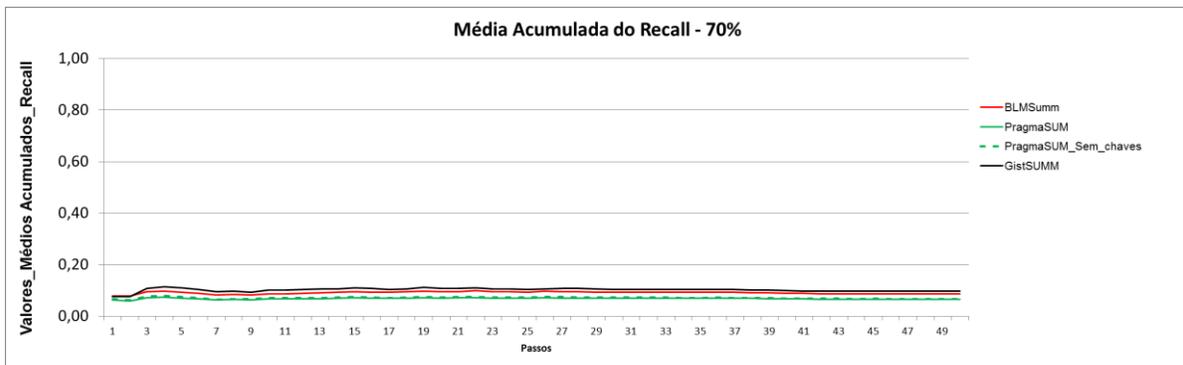


Figura 38. Média acumulada de Recall com taxa de compressão de 70% do domínio Administração.

1.3. Taxa de Compressão de 80%

Na Figura 39 (média acumulada de F-Measure) percebe-se que o BLMSumm e o GistSumm apresentam resultados muito próximos e tem pouca diferença para o PragmaSUM.

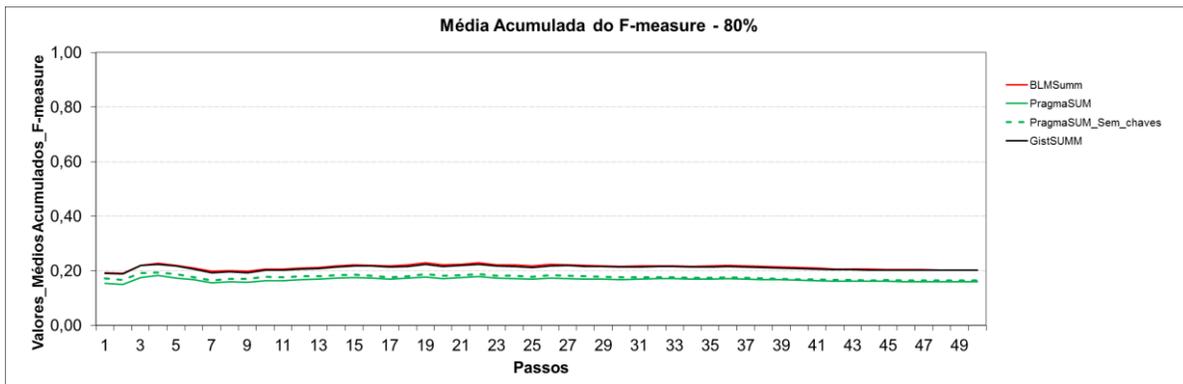


Figura 39. Média acumulada de F-Measure com taxa de compressão de 80% do domínio Administração.

A Figura 40 (média acumulada de Precision) mostram que, de uma maneira geral, as taxas de precisão alcançadas foram altas e que o sumarizadores que obtiveram os melhores valores foram as duas versões do PragmaSUM seguidos por BLMSumm e GistSumm com valores muito próximos.

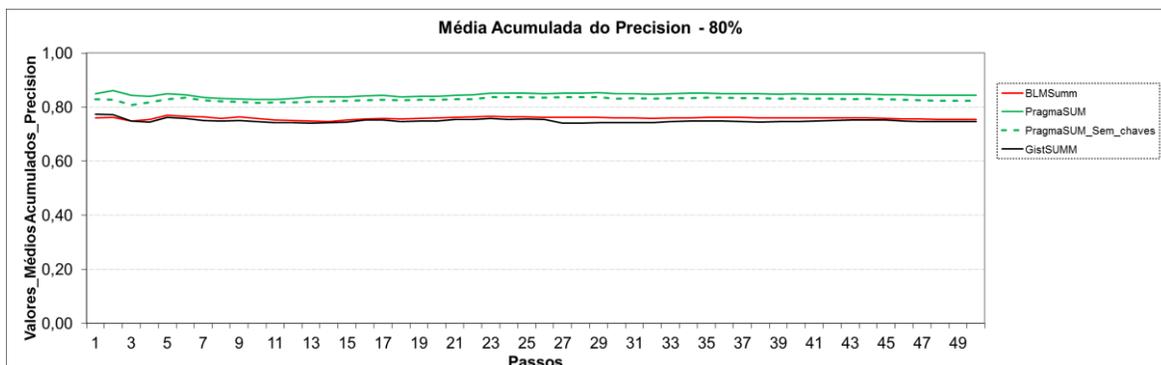


Figura 40. Média acumulada de Precision com taxa de compressão de 80% do domínio Administração.

A Figura 41 (média acumulada de Recall) evidenciam que quanto à cobertura os resultados foram baixos e os sumarizadores tiveram valores bem próximos, com uma ligeira vantagem para o BLMSumm e para o GistSumm.

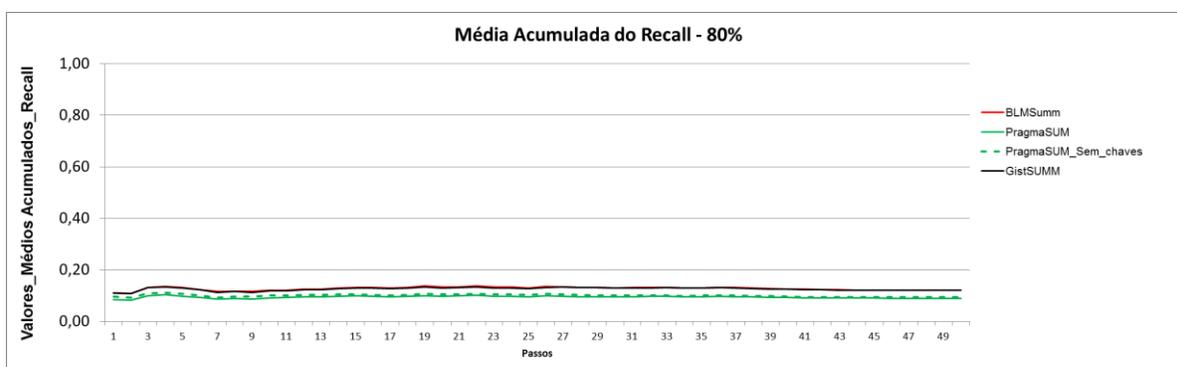


Figura 41. Média acumulada de Recall com taxa de compressão de 80% do domínio Administração.

1.4. Taxa de Compressão de 90%

É possível perceber na Figura 42 (média acumulada de F-Measure) que, de forma geral, o BLMSumm e o GistSumm apresenta resultados bem aproximados e os mais satisfatórios.

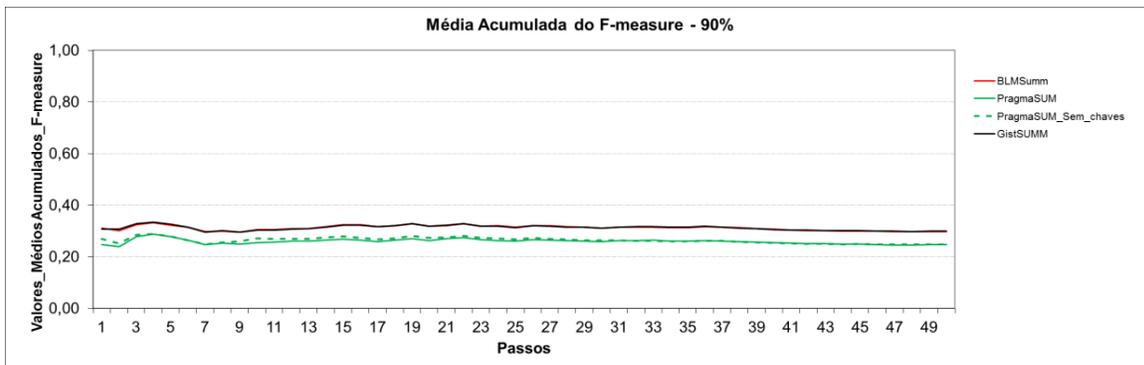


Figura 42. Média acumulada de F-Measure com taxa de compressão de 90% do domínio Administração.

É possível perceber na Figura 43 (média acumulada de Precision) que quanto à precisão o PragmaSUM obteve os melhores resultados com uma ligeira vantagem sobre sua versão sem chaves. Os resultados mais baixos foram obtidos pelo BLMSumm e GistSumm e obtiveram valores próximos entre si.

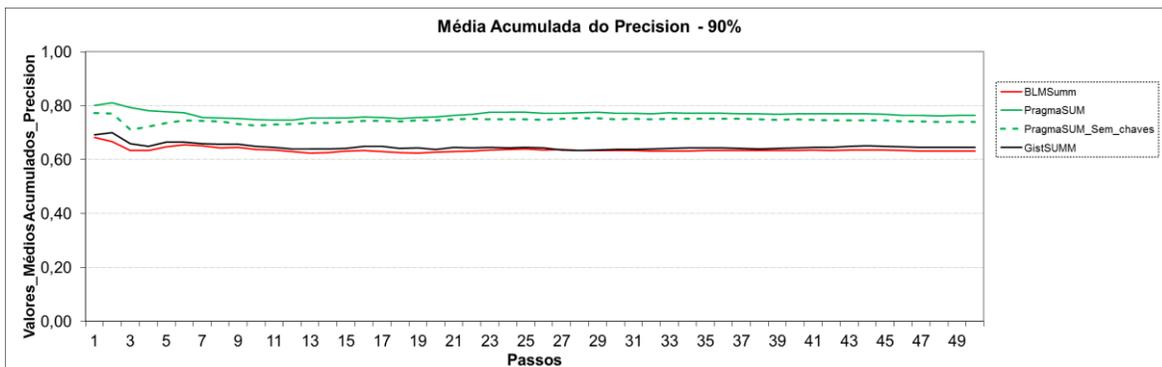


Figura 43. Média acumulada de Precision com taxa de compressão de 90% do domínio Administração.

É possível perceber Figura 44 (média acumulada de Recall) que quanto à cobertura o BLMSumm e o GistSumm obtiveram os melhores resultados com valores muito próximos.

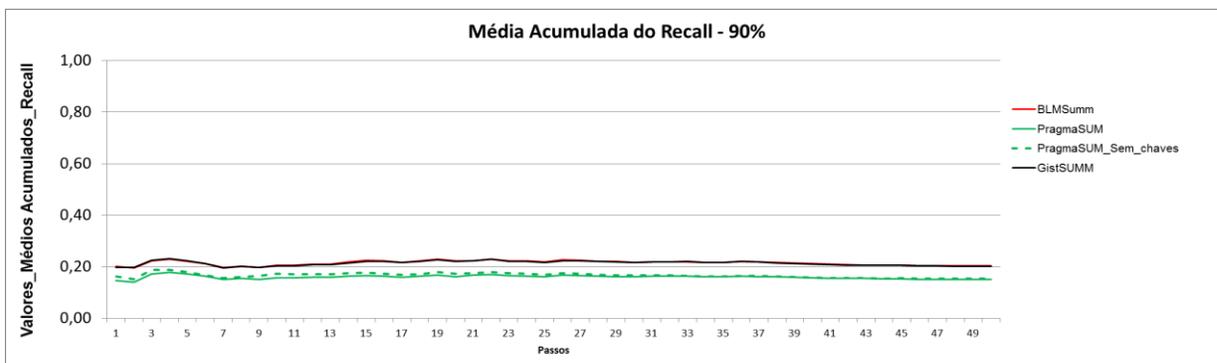


Figura 44. Média acumulada de Recall com taxa de compressão de 90% do domínio Administração

1.5. Comparação dos Resultados

A Tabela 17 mostra uma comparação feita entre as taxas de compressão das médias gerais de média harmônica (F-Measure) alcançadas pelos sumarizadores automáticos para os textos do domínio Administração. O GistSumm obteve os melhores resultados nas compressões de 50 e 70%, com 80 e 90% os valores entre GistSumm e BLMSumm foram bem aproximados. As duas versões do PragmaSUM tiveram os valores mais baixos e quase idênticos. Quando a taxa de compressão aumenta os valores ficam mais homogêneos. A Figura 45 reafirma as observações.

Tabela 17. Comparação dos resultados de F-Measure do domínio Administração.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,10357	0,15867	0,0841	0,08541
70%	0,15311	0,16752	0,12017	0,1235
80%	0,20261	0,20153	0,15959	0,16485
90%	0,29702	0,29819	0,24682	0,24732

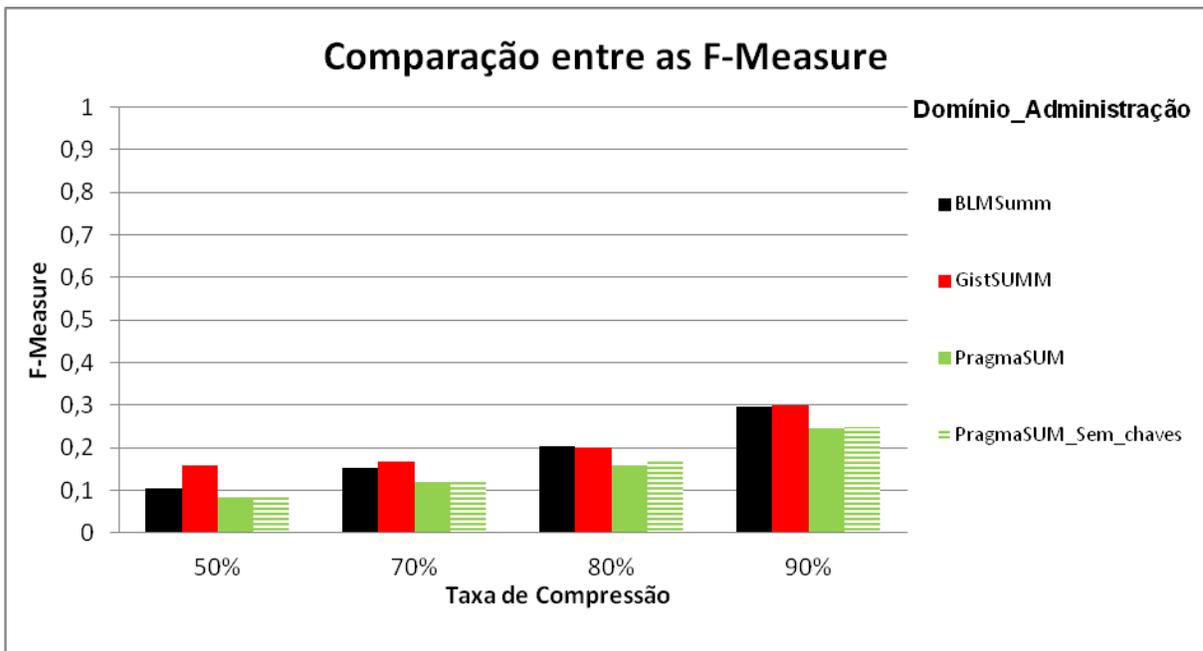


Figura 45. Comparação dos resultados de F-Measure do domínio Administração.

A Tabela 18 e a Figura 46 trazem uma comparação feita entre as taxas de compressão das médias gerais de Recall (cobertura) que os sumarizadores automáticos alcançaram para os textos do domínio Administração. Constata-se que o GistSumm apresentou os melhores resultados para as 2 primeiras taxas de compressão aplicadas e para as duas últimas os resultados ficaram muito próximos com o BLMSumm. As duas versões do PragmaSUM obtiveram valores muito próximos.

Tabela 18. Comparação dos resultados de Recall do domínio Administração.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,05557	0,09142	0,0444	0,04517
70%	0,08588	0,09677	0,0652	0,06742
80%	0,11995	0,11983	0,08951	0,09344
90%	0,20264	0,20208	0,1514	0,15357

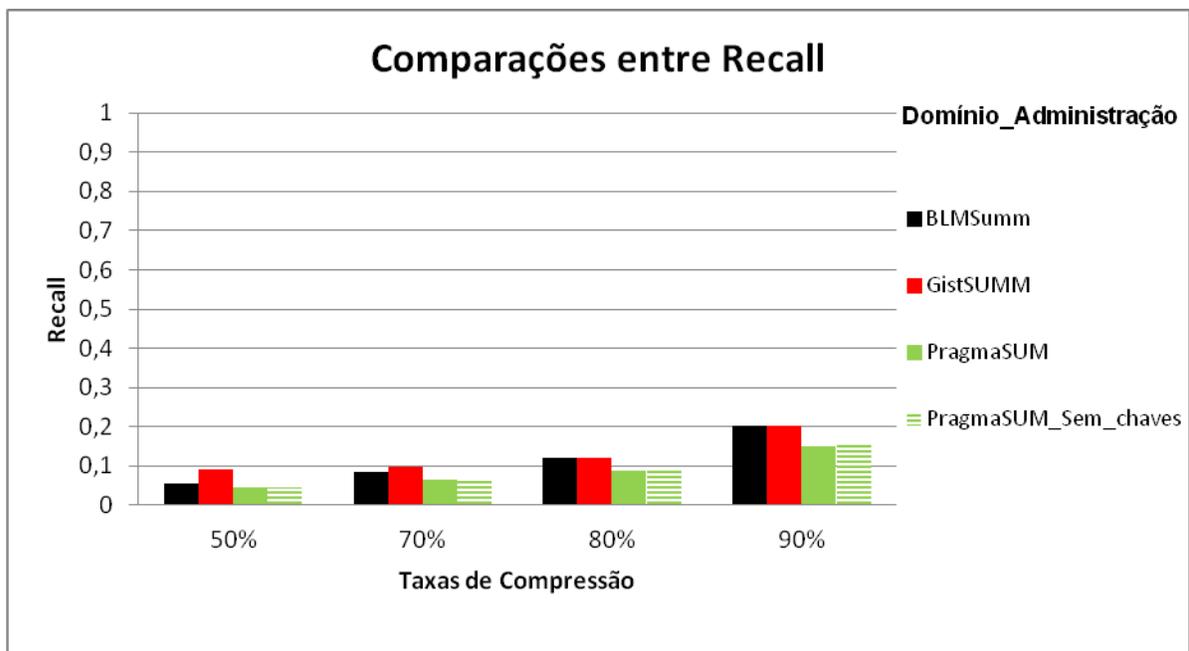


Figura 46. Comparação dos resultados de Recall do domínio Administração.

2. Domínio Agronomia

Esta seção descreve os resultados das SA realizadas no domínio Agronomia.

2.1. Taxa de Compressão de 50%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio Agronomia. A taxa de compressão aplicada foi de 50%.

A figura a seguir (média acumulada de F-Measure) apresentam os resultados das SA para cada um dos sumarizadores automáticos. Observa-se que o GistSumm possui uma ligeira vantagem.

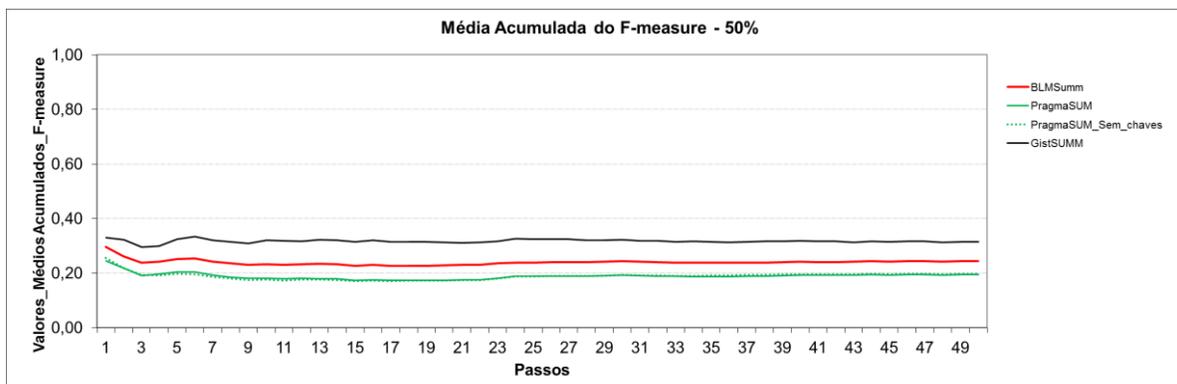


Figura 47. Média acumulada de F-Measure com taxa de compressão de 50% do domínio Agronomia.

Na figura a seguir (média acumulada de Precision) têm-se os resultados para cada um dos sumarizadores automáticos. Quanto à precisão (Precision) as duas versões do PragmaSUM apresentaram os melhores resultados, e muito próximos entre si, o GistSumm apresentou os valores mais baixos.

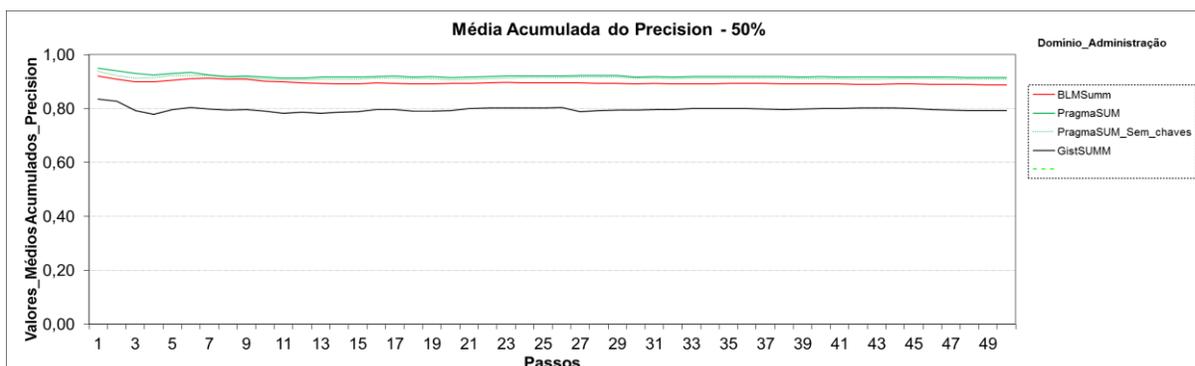


Figura 48. Média acumulada de Precision com taxa de compressão de 50% do domínio Agronomia.

A figura a seguir (média acumulada de Recall) mostram os resultados da cobertura para cada um dos sumarizadores automáticos. É possível perceber que existe uma variação muito pequena dentre resultados aferidos e o GistSumm possui ligeira vantagem.

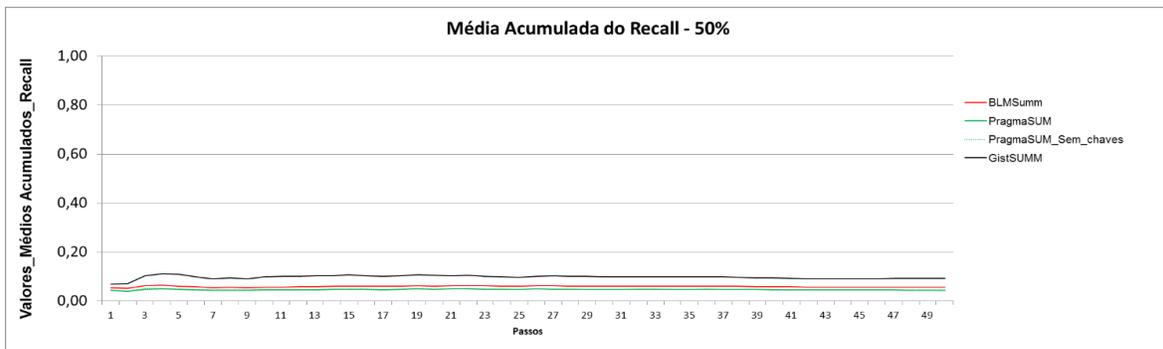


Figura 49. Média acumulada de Recall com taxa de compressão de 50% do domínio Agronomia.

2.2. Taxa de Compressão de 70%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio Agronomia para uma taxa de compressão de 70%.

É possível perceber na figura seguir (média acumulada de F-Measure) que o GistSumm apresenta uma ligeira vantagem sobre o BLMSumm. Nota-se ainda que os valores são bem próximos.

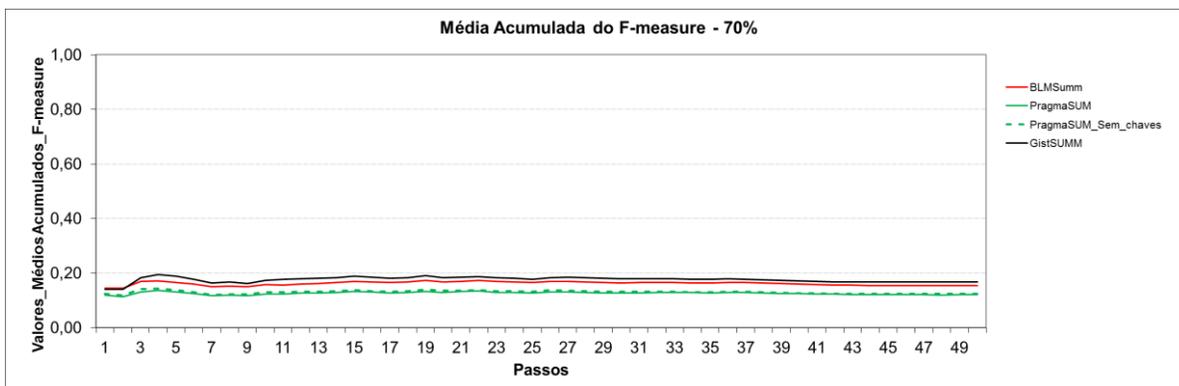


Figura 50. Média acumulada de F-Measure com taxa de compressão de 70% do domínio Agronomia.

Percebe-se na figura a seguir (média acumulada de Precision) que o sumarizador com os valores de precisão mais altos foi o PragmaSUM, seguido pelo sua versão sem chaves e o com valores mais baixos foi o GistSumm.

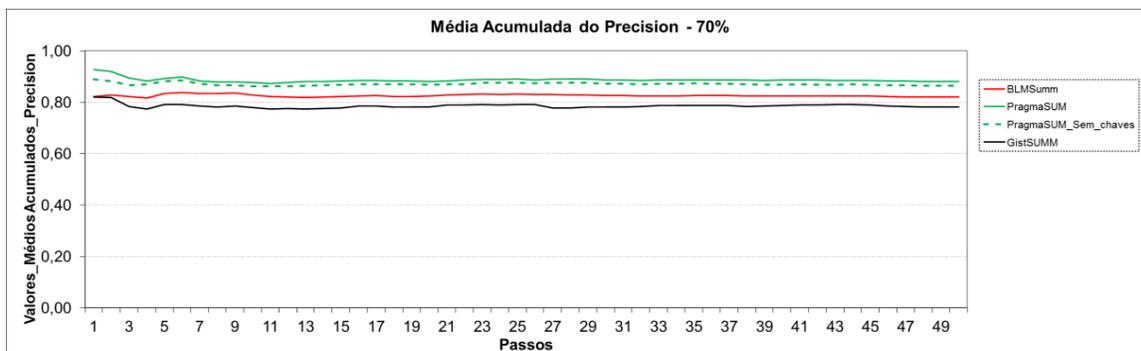


Figura 51. Média acumulada de Precision com taxa de compressão de 70% do domínio Agronomia.

A tabela 33 e a figura 33 (média acumulada de Recall) mostram que quanto à cobertura os sumarizadores que obtiveram resultados muito próximos e valores muito baixos.

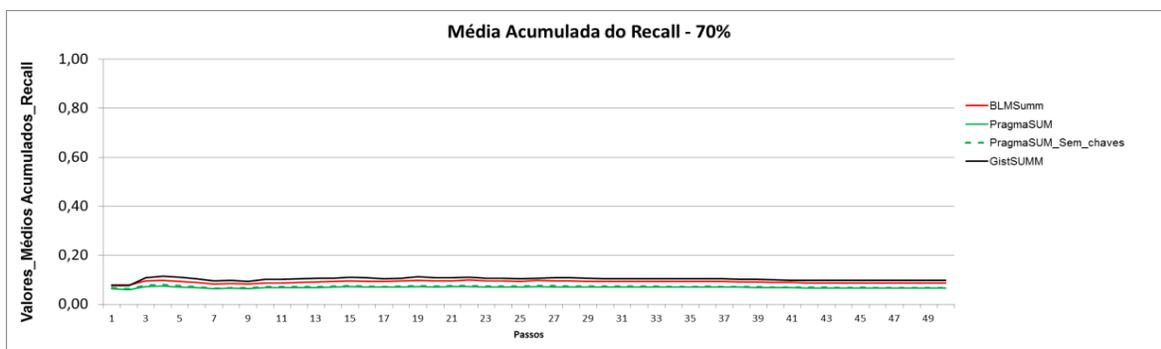


Figura 52. Média acumulada de Recall com taxa de compressão de 70% do domínio Agronomia.

2.3. Taxa de Compressão de 80%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio Agronomia com taxa de compressão de 80%.

Na figura a seguir (média acumulada de F-Measure) percebe-se que o BLMSumm e o GistSumm apresentam os resultados mais satisfatórios mas com pouca diferença sobre o PragmaSUM.

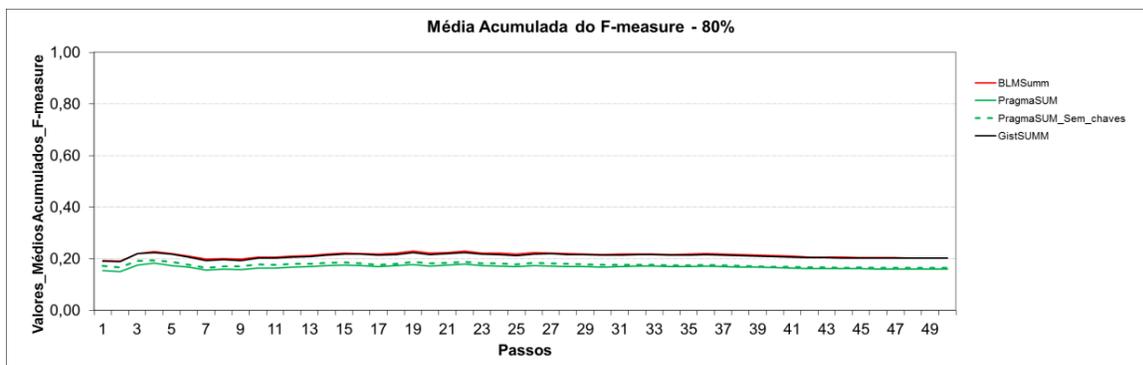


Figura 53. Média acumulada de F-Measure com taxa de compressão de 80% do domínio Agronomia.

A figura a seguir (média acumulada de Precision) mostram que, de uma maneira geral, as taxas de precisão alcançadas foram altas e que o sumariador que obteve os melhores valores foram as duas versões do PragmaSUM, com ligeira vantagem sobre sua versão que utiliza as palavras-chave.

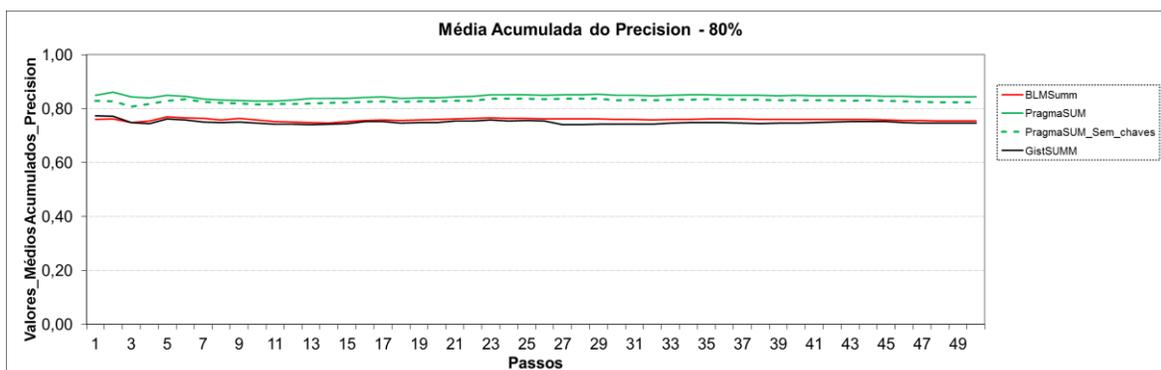


Figura 54. Média acumulada de Precision com taxa de compressão de 80% do domínio Agronomia.

A figura a seguir (média acumulada de Recall) evidenciam que quanto à cobertura os sumariadores obtiveram resultados muito próximos.

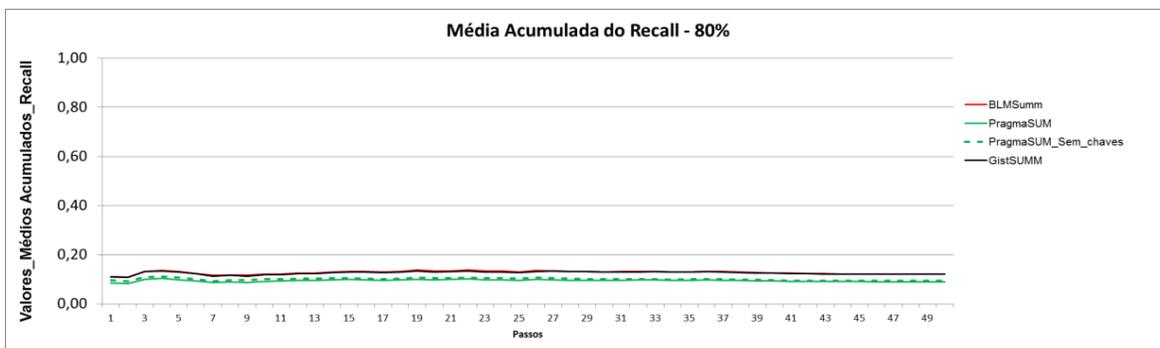


Figura 55. Média acumulada de Recall com taxa de compressão de 80% do domínio agronomia.

2.4. Taxa de Compressão de 90%

Nas figura desta seção têm-se os resultados das SA para os 50 textos do domínio agronomia com taxa de compressão aplicada de 90%.

É possível perceber na figura a seguir (média acumulada de F-Measure) que, de forma geral, o BLMSumm junto com o GistSumm apresenta os resultados mais satisfatórios.

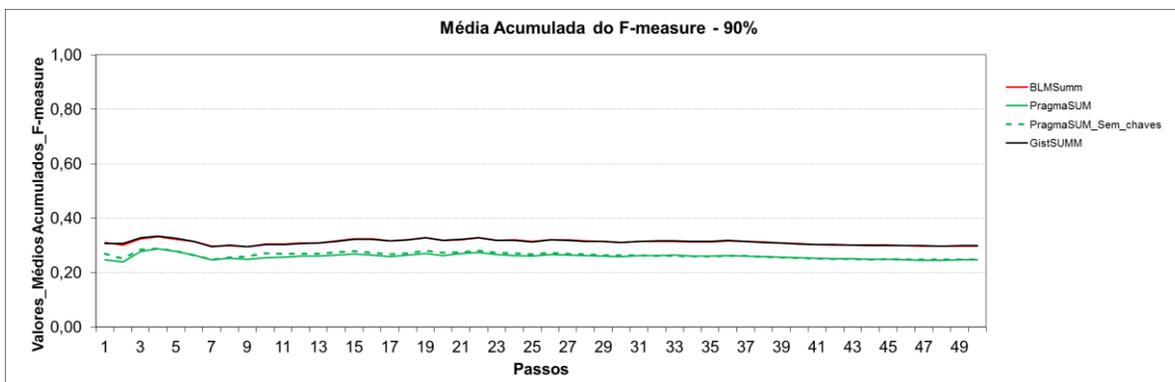


Figura 56. Média acumulada de F-Measure com taxa de compressão de 90% do domínio Agronomia.

É possível na figura a seguir (média acumulada de Precision) que quanto à precisão o PragmaSUM obteve os melhores resultados dentre os sumarizadores analisados e que estes valores foram muito bons, com ligeira vantagem sobre sua versão sem chaves. Os resultados mais baixos foram obtidos pelo BLMSumm e GistSumm com valores bem próximos entre si.

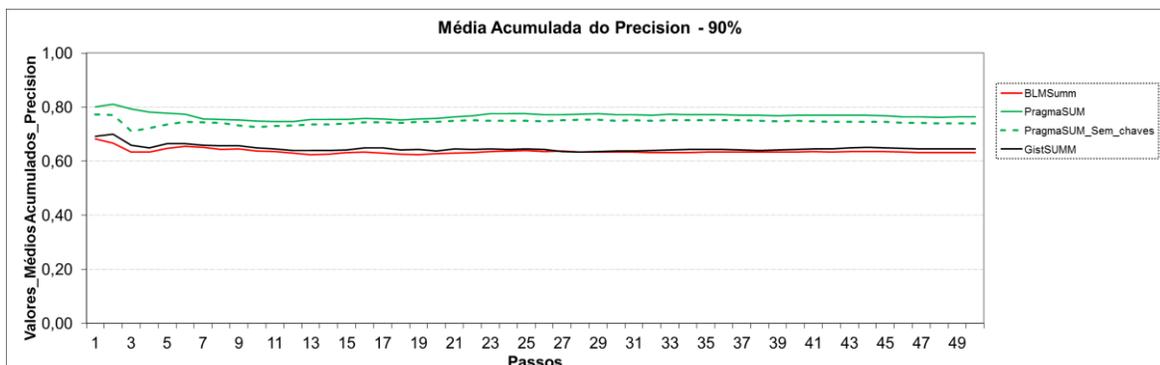


Figura 57. Média acumulada de Precision com taxa de compressão de 90% do domínio Agronomia.

É possível perceber na figura a seguir (média acumulada de Recall) que quanto à cobertura o BLMSumm obteve os melhores resultados junto com o GistSumm.

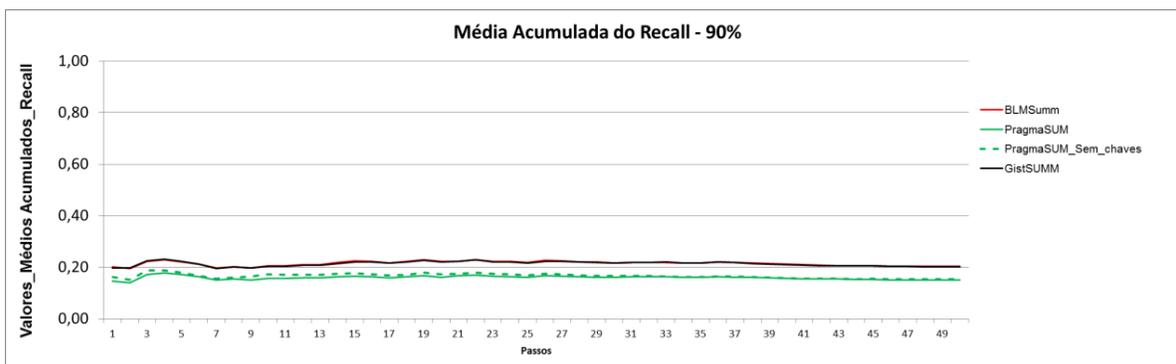


Figura 58. Média acumulada de Recall com taxa de compressão de 90% do domínio Agronomia.

2.5. Comparação dos Resultados

A tabela a seguir mostra uma comparação feita entre as taxas de compressão das médias gerais de média harmônica (F-Measure) alcançadas pelos sumarizadores automáticos para os textos do domínio Agronomia. O GistSumm obteve os melhores resultados na taxa de 50%, nas taxas de 70, 80 e 90% o BLMSumm se igualou ao GistSumm com valores bem aproximados. O PragmaSUM teve valores um pouco menores e bem próximos em suas duas versões. A Figura 59 reafirma as observações.

Tabela 19. Comparação dos resultados de F-Measure do domínio Agronomia.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,24347	0,31408	0,19458	0,19702
70%	0,3211	0,32286	0,25297	0,26528
80%	0,37838	0,35833	0,30593	0,31793
90%	0,42329	0,40854	0,39453	0,38275

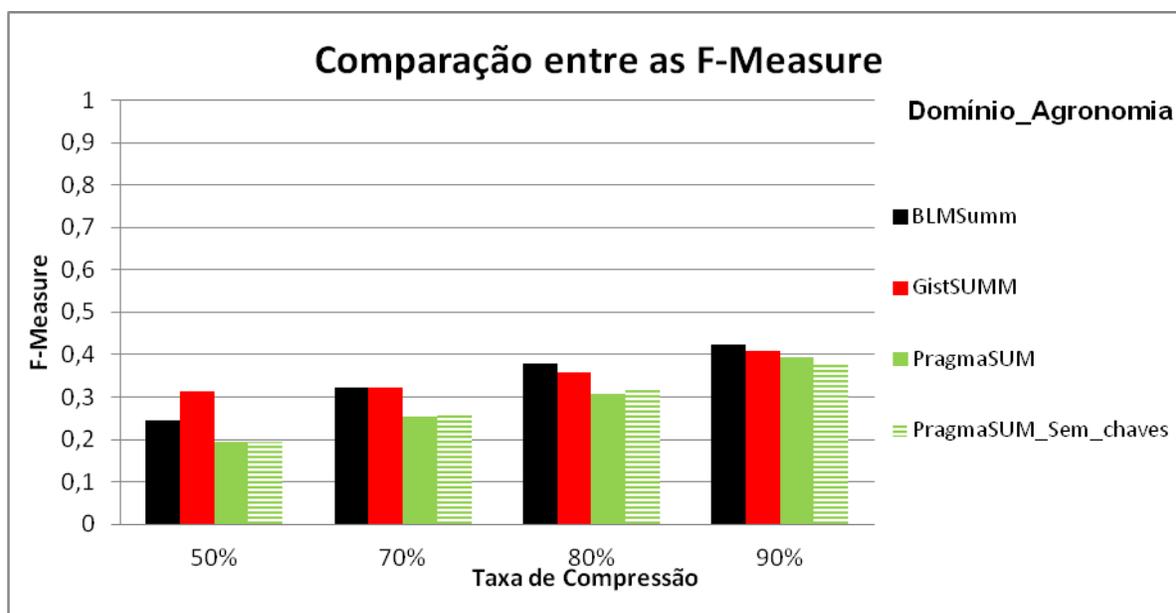


Figura 59. Comparação dos resultados de F-Measure do domínio agronomia.

A tabela e a figura a seguir trazem uma comparação feita entre as taxas de compressão das médias gerais de Recall (cobertura) que os sumarizadores automáticos alcançaram para os textos do domínio agronomia. Constata-se que o BLMSumm e o GistSumm apresentaram os melhores resultados para as 4 taxas de compressão aplicadas.

Tabela 20. Comparação dos resultados de Recall do domínio Agronomia.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,14489	0,2216	0,11111	0,11282
70%	0,21121	0,22994	0,15344	0,16344
80%	0,27969	0,26887	0,19802	0,2126
90%	0,40617	0,39417	0,30328	0,31887

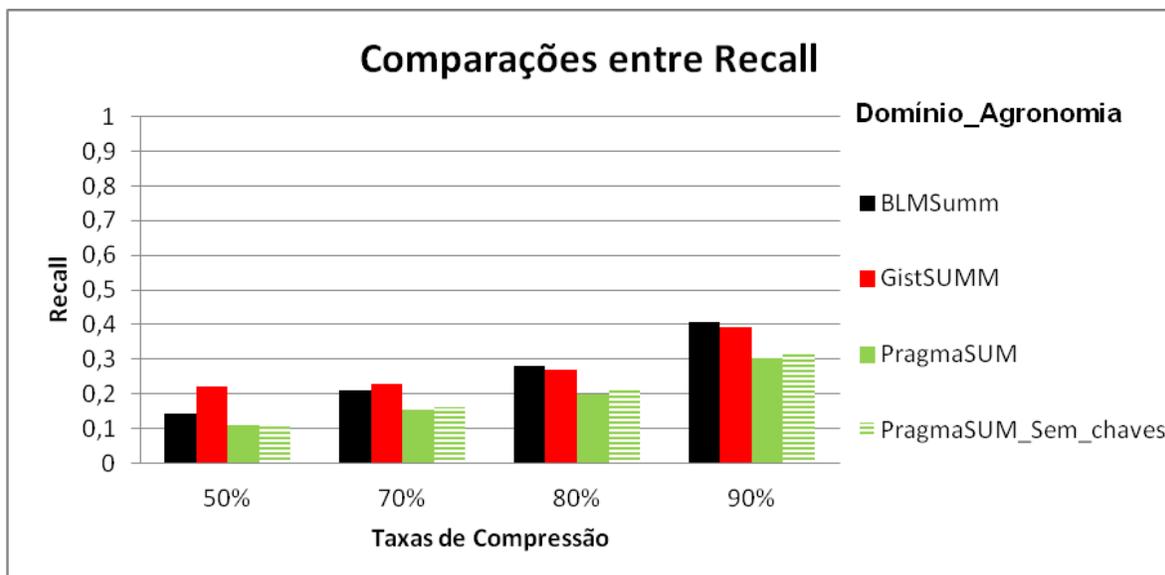


Figura 60. Comparação dos resultados de Recall do domínio agronomia.

3. Domínio Educação Física

Esta seção descreve os resultados das SA realizadas no domínio Educação Física.

3.1. Taxa de Compressão de 50%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio Educação Física. A taxa de compressão aplicada foi de 50%.

A figura a seguir (média acumulada de F-Measure) apresentam os resultados das SA para cada um dos sumarizadores automáticos. Observa-se que há bastante equilíbrio nos resultados, mas o GistSumm possui vantagem.

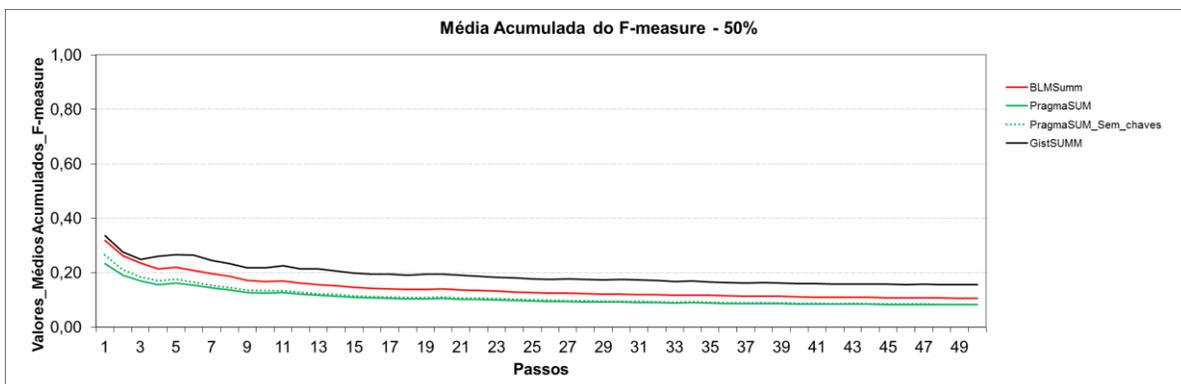


Figura 61. Média acumulada de F-Measure com taxa de compressão de 50% do domínio Educação Física.

Na figura a seguir (média acumulada de Precision) têm-se os resultados para cada um dos sumarizadores automáticos. Quanto à precisão (Precision) o PragmaSUM apresentou os melhores resultados.

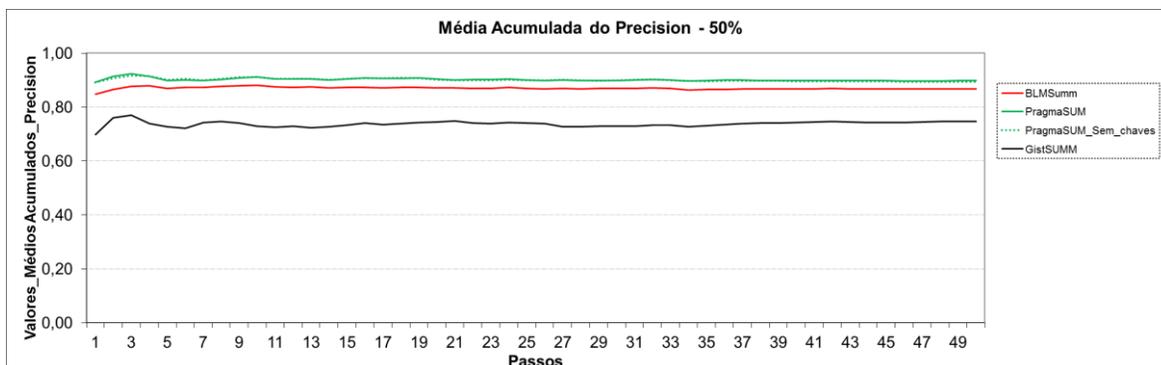


Figura 62. Média acumulada de Precision com taxa de compressão de 50% do domínio Educação Física.

A figura a seguir (média acumulada de Recall) mostram os resultados da cobertura para cada um dos sumarizadores automáticos. É possível perceber que existe uma variação muita pequena dentre resultados aferidos e o GistSumm possui ligeira vantagem.

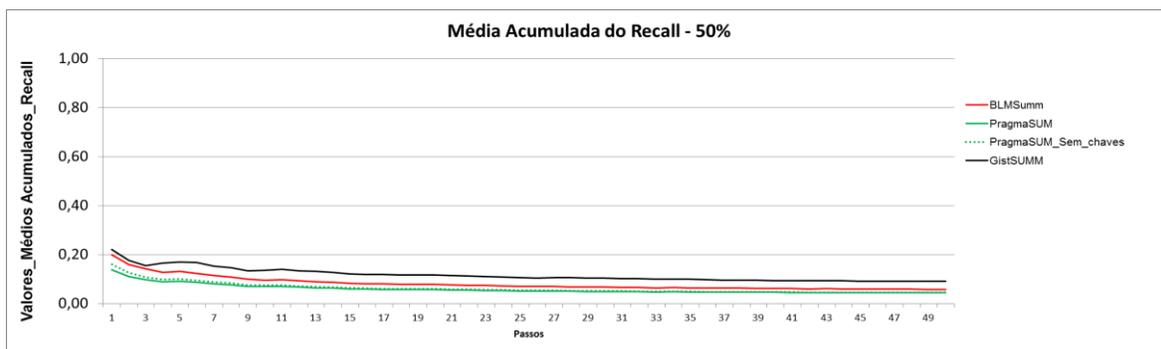


Figura 45. Média acumulada de Recall com taxa de compressão de 50% do domínio Educação Física.

3.2. Taxa de Compressão de 70%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio Educação Física para uma taxa de compressão de 70%.

É possível perceber na figura a seguir (média acumulada de F-Measure) que existe uma variação muito pequena dentre resultados aferidos e o GistSumm possui ligeira vantagem.

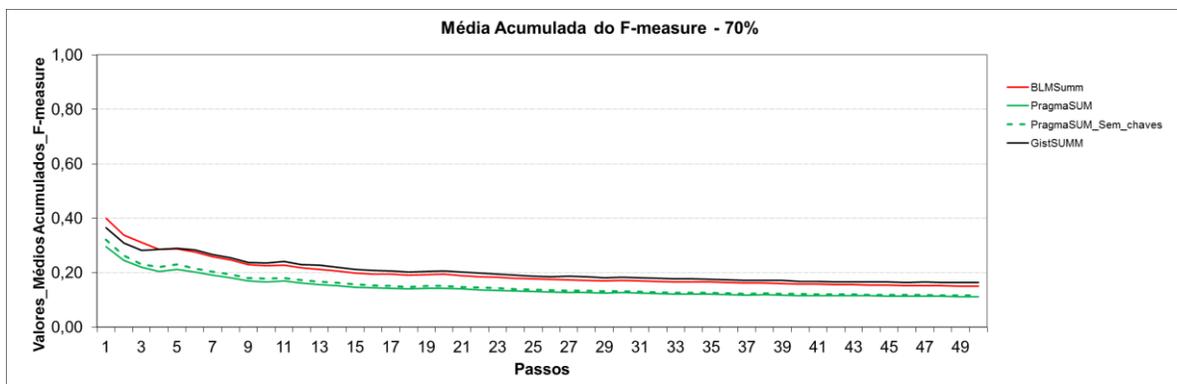


Figura 63. Média acumulada de F-Measure com taxa de compressão de 70% do domínio Educação Física.

Percebe-se na figura a seguir (média acumulada de Precision) que o sumarizador com os valores de precisão mais altos foi o PragmaSUM, seguido pelo BLMSum e com valores mais baixos foi o GistSumm.

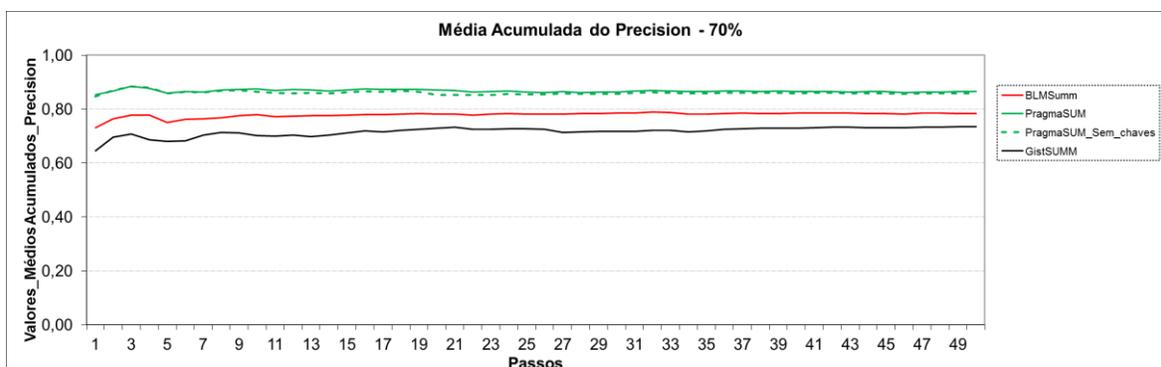


Figura 64. Média acumulada de Precision com taxa de compressão de 70% do domínio Educação Física.

A figura a seguir (média acumulada de Recall) mostram que quanto à cobertura os sumarizadores que obtiveram ligeira vantagem foram o BLMSum e o GistSumm.

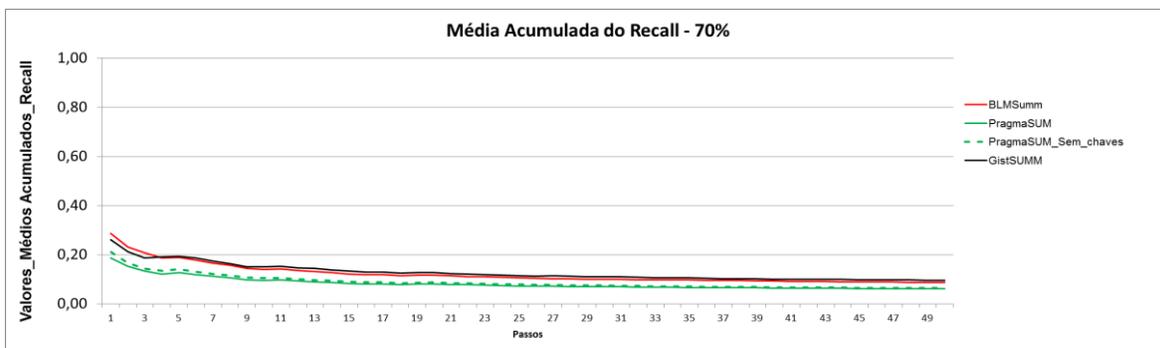


Figura 65. Média acumulada de Recall com taxa de compressão de 70% do domínio Educação Física.

3.3. Taxa de Compressão de 80%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio Educação Física com taxa de compressão de 80%.

Na figura a seguir (média acumulada de F-Measure) percebe-se que os sumarizadores que obtiveram ligeira vantagem foram o BLMSumm e o GistSumm.

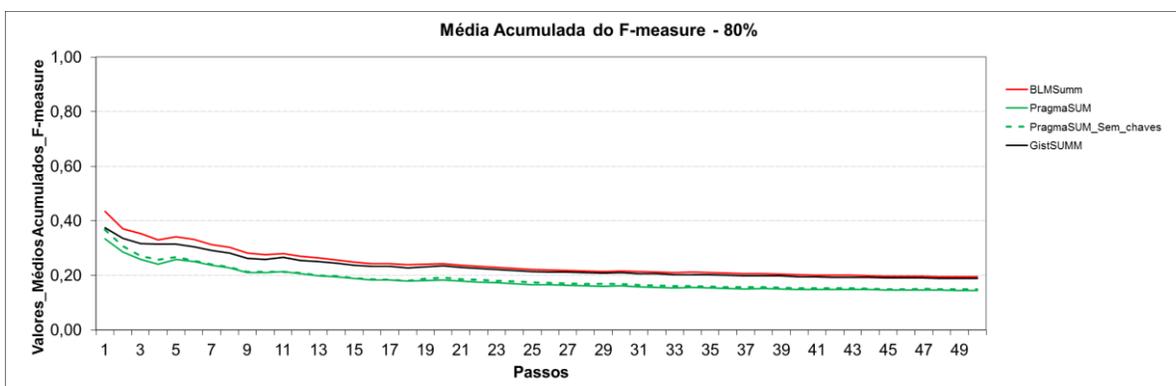


Figura 66. Média acumulada de F-Measure com taxa de compressão de 80% do domínio Educação Física.

A figura a seguir (média acumulada de Precision) mostram que, de uma maneira geral, as taxas de precisão alcançadas foram altas e que o sumarizadores que obtiveram os melhores valores foram as duas versões do PragmaSUM.

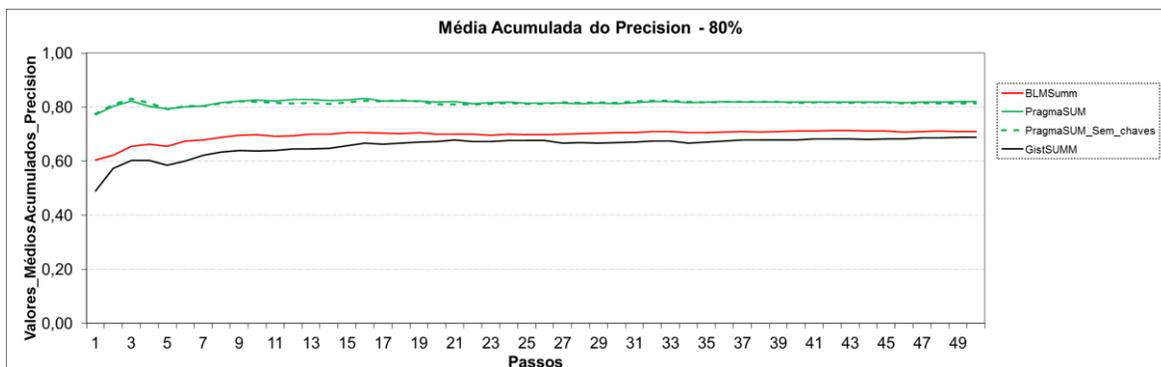


Figura 67. Média acumulada de Precision com taxa de compressão de 80% do domínio Educação Física.

A figura a seguir (média acumulada de Recall) evidenciam que os sumarizadores que obtiveram ligeira vantagem foram o BLMSumm e o GistSumm.

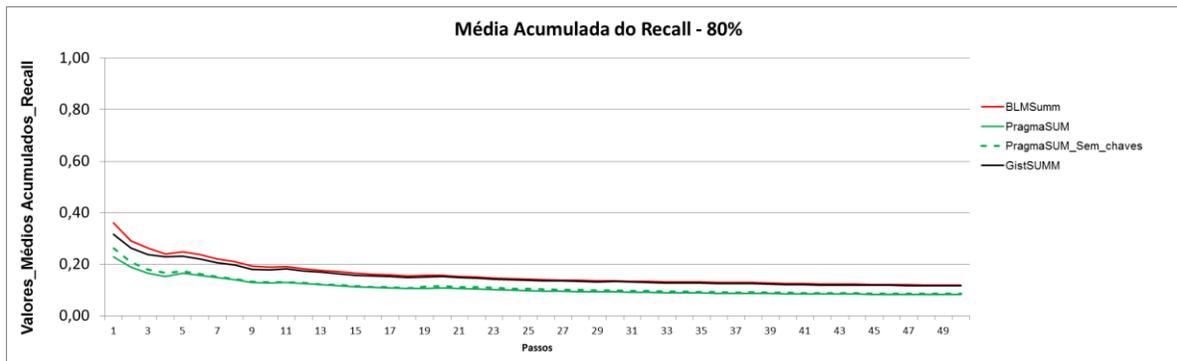


Figura 68. Média acumulada de Recall com taxa de compressão de 80% do domínio Educação Física.

3.4. Taxa de Compressão de 90%

Nas figuras desta seção têm-se os resultados das SA para os 50 textos do domínio Educação Física com taxa de compressão aplicada de 90%.

É possível perceber na figura a seguir (média acumulada de F-Measure) que, de forma geral, o GistSumm apresenta os ligeira vantagem sobre os demais.

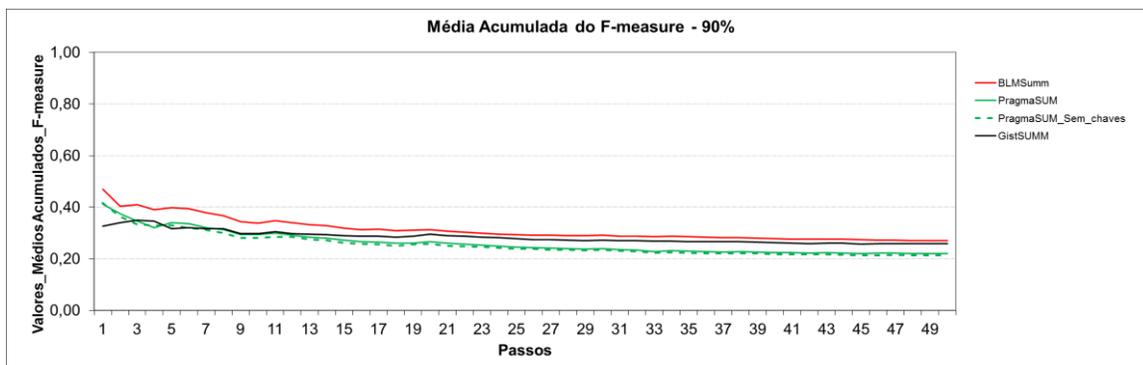


Figura 69. Média acumulada de F-Measure com taxa de compressão de 90% do domínio Educação Física.

É possível perceber na figura a seguir (média acumulada de Precision) que quanto à precisão o PragmaSUM obteve os melhores resultados com ligeira vantagem sobre sua versão sem chaves. BLMSumm e GistSumm obtiveram valores próximos entre si.

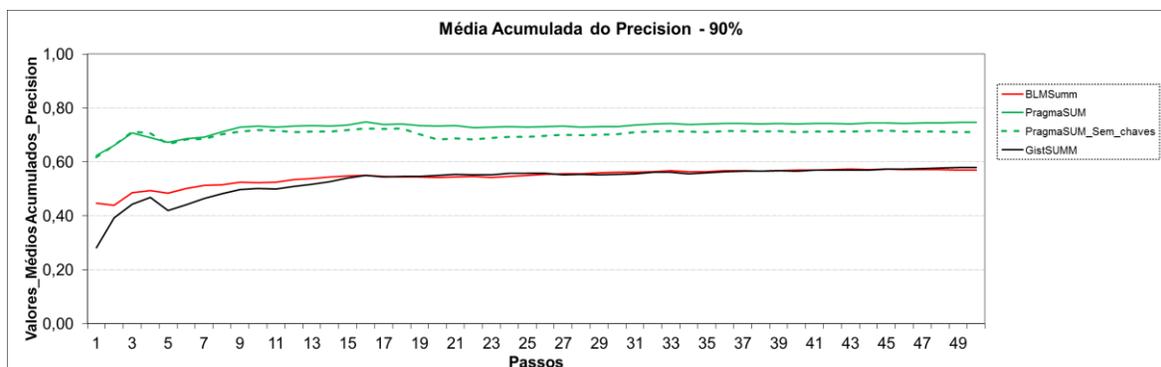


Figura 70. Média acumulada de Precision com taxa de compressão de 90% do domínio Educação Física.

É possível perceber na figura a seguir (média acumulada de Recall) que quanto à cobertura o GistSumm obteve os melhores resultados com ligeira vantagem sobre os demais.

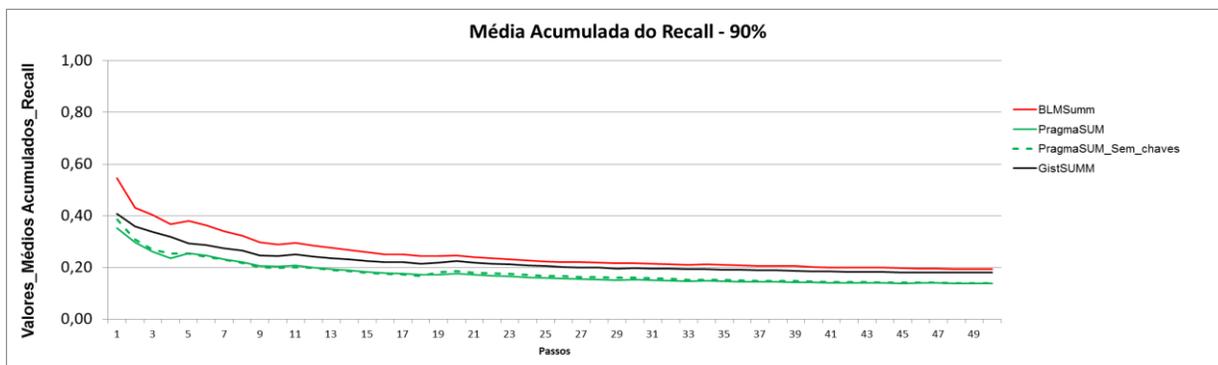


Figura 71. Média acumulada de Recall com taxa de compressão de 90% do domínio Educação Física.

3.5. Comparação dos Resultados

A tabela a seguir mostra uma comparação feita entre as taxas de compressão das médias gerais de média harmônica (F-Measure) alcançadas pelos cinco sumarizadores automáticos para os textos do domínio agronomia. O BLMSumm e o GistSumm obtiveram os melhores resultados. Os valores variam pouco entre si. A Figura 72 reafirma as observações.

Tabela 21. Comparação dos resultados de F-Measure do domínio Educação Física.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,10605	0,15514	0,08157	0,08347
70%	0,15071	0,16266	0,11199	0,11634
80%	0,19394	0,18859	0,14482	0,14811
90%	0,26947	0,2577	0,21994	0,21181

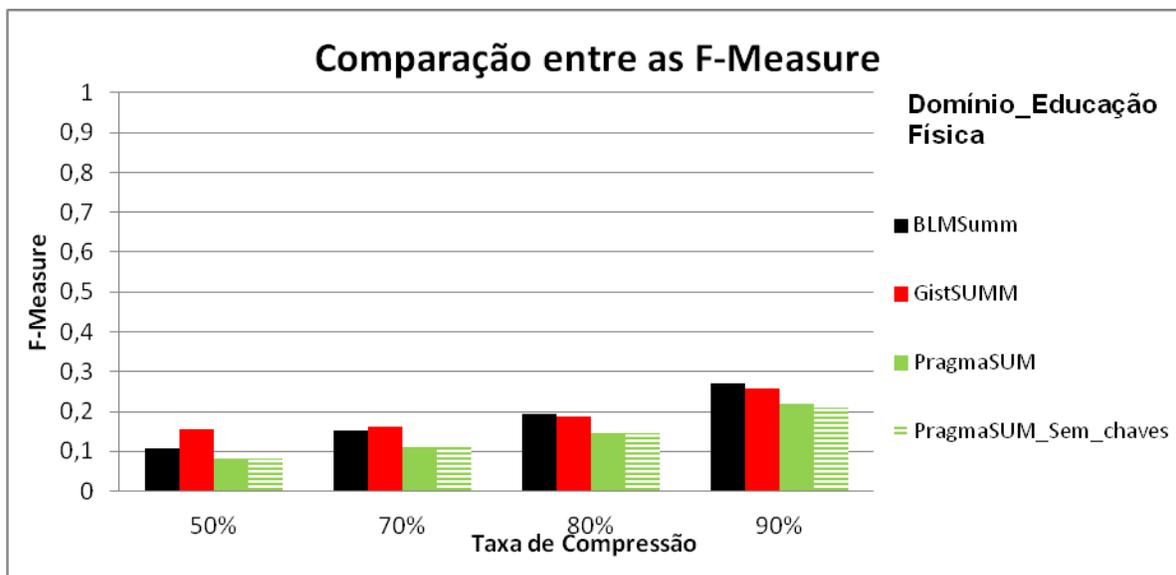


Figura 72. Comparação dos resultados de F-Measure do domínio Educação Física.

A tabela e a figura a seguir trazem uma comparação feita entre as taxas de compressão das médias gerais de Recall (cobertura) que os cinco sumarizadores automáticos alcançaram para os textos do domínio Agronomia. Constata-se que o BLMSumm e o GistSumm apresentaram os melhores resultados para as 4 taxas de compressão aplicadas com ligeira vantagem.

Tabela 22. Comparação dos resultados de Recall do domínio Educação Física.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,05805	0,09046	0,04358	0,04484
70%	0,08694	0,09598	0,06154	0,06456
80%	0,11869	0,1158	0,08248	0,08531
90%	0,19217	0,17885	0,13731	0,13801

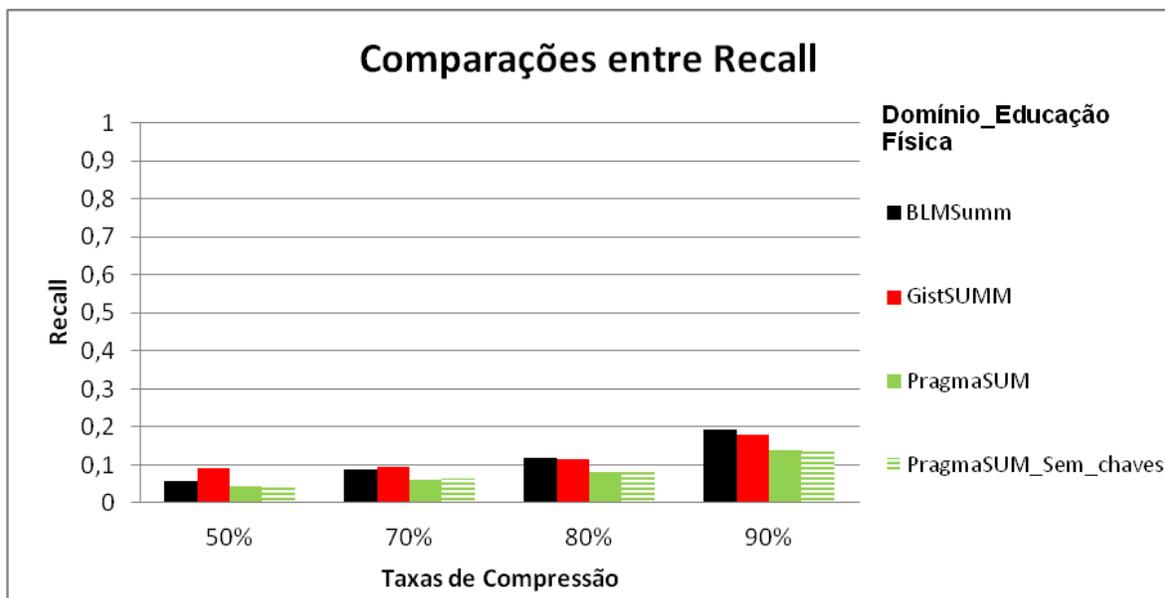


Figura 73. Comparação dos resultados de Recall do domínio Educação Física.

4. Domínio Engenharia

Esta seção descreve os resultados das SA realizadas no domínio Engenharia.

4.1. Taxa de Compressão de 50%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio Engenharia. A taxa de compressão aplicada foi de 50%.

A figura a seguir (média acumulada de F-Measure) apresentam os resultados das SA para cada um dos sumarizadores automáticos. Observa-se que há bastante equilíbrio nos resultados, mas o GistSumm possui os melhores resultados.

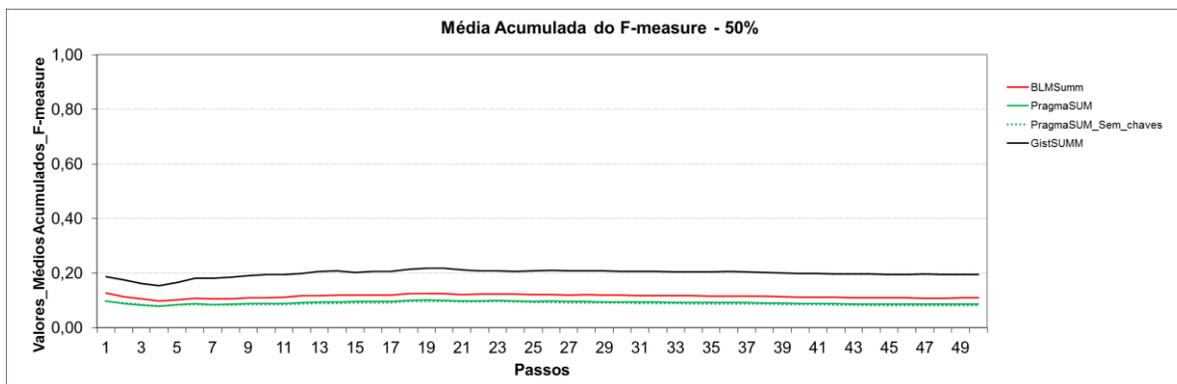


Figura 74. Média acumulada de F-Measure com taxa de compressão de 50% do domínio Engenharia.

Na figura a seguir (média acumulada de Precision) têm-se os resultados para cada um dos sumarizadores automáticos. Quanto à precisão (Precision) o PragmaSUM apresentou os melhores resultados.

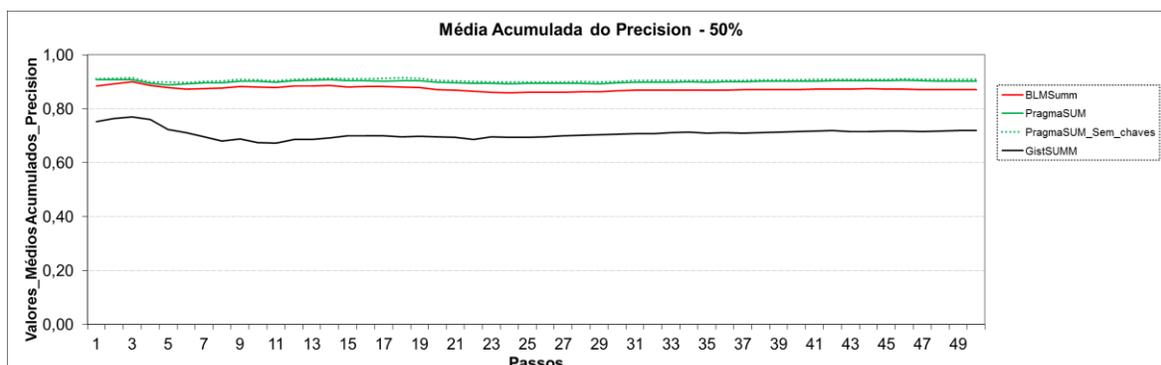


Figura 75. Média acumulada de Precision com taxa de compressão de 50% do domínio Engenharia.

A figura a seguir (média acumulada de Recall) mostram os resultados da cobertura para cada um dos sumarizadores automáticos. É possível perceber que existe uma variação muita pequena dentre resultados aferidos e o GistSumm possui ligeira vantagem.

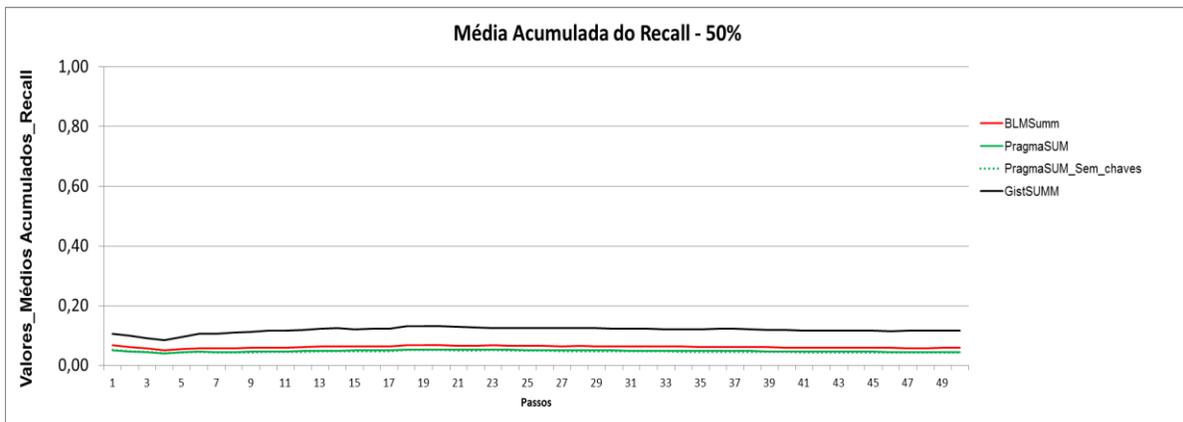


Figura 76. Média acumulada de Recall com taxa de compressão de 50% do domínio Engenharia.

4.2. Taxa de Compressão de 70%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio Engenharia para uma taxa de compressão de 70%.

É possível perceber figura a seguir (média acumulada de F-Measure) que o GistSumm apresenta ligeira vantagem sobre os demais.

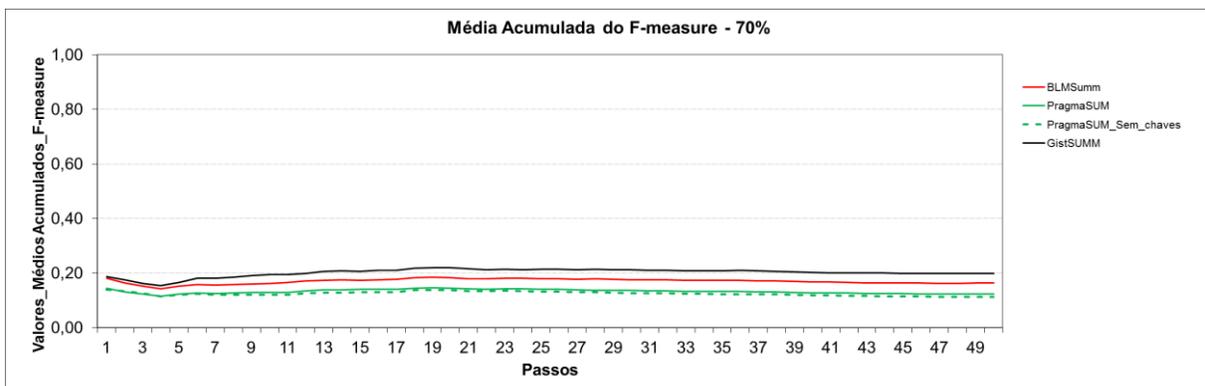


Figura 77. Média acumulada de F-Measure com taxa de compressão de 70% do domínio Engenharia.

Percebe-se na figura a seguir (média acumulada de Precision) que o sumarizador com os valores de precisão mais altos foi o PragmaSUM, seguido pelo BLMSumm e o com valores mais baixos foi o GistSumm.

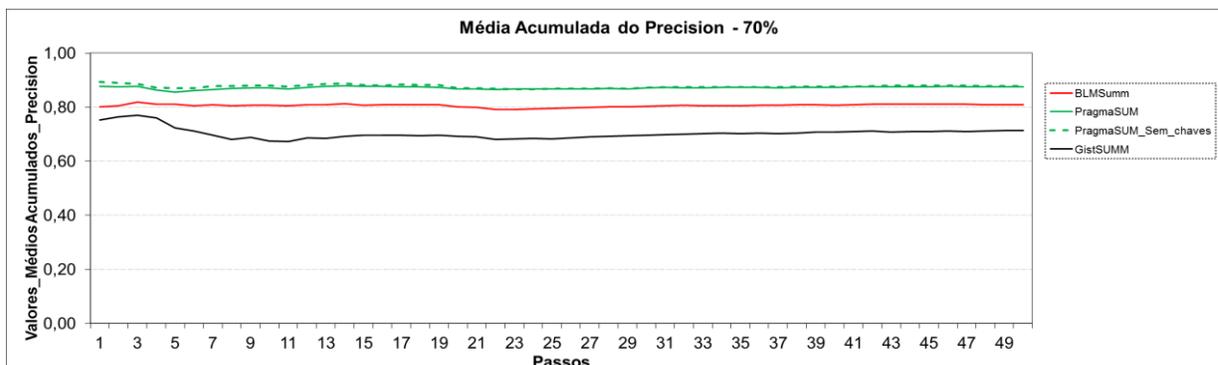


Figura 78. Média acumulada de Precision com taxa de compressão de 70% do domínio Engenharia.

A figura a seguir (média acumulada de Recall) mostram que quanto à cobertura os sumarizadores que obtiveram os resultados muito próximos.

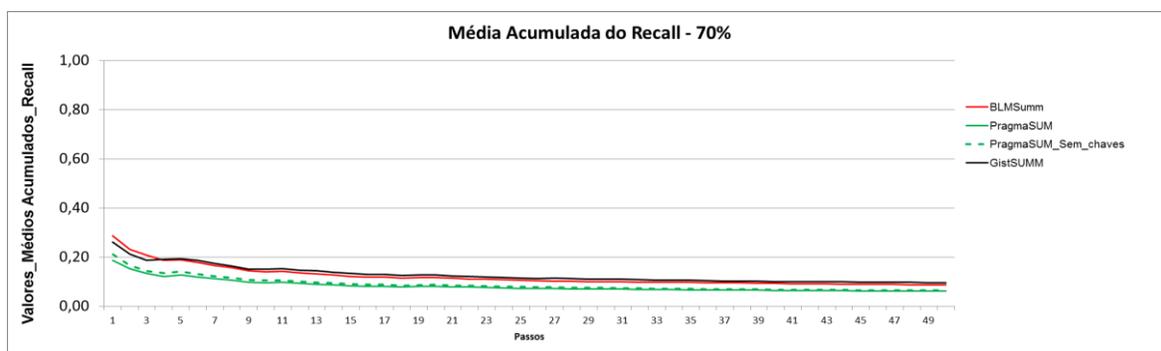


Figura 79. Média acumulada de Recall com taxa de compressão de 70% do domínio Engenharia.

4.3. Taxa de Compressão de 80%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio Engenharia com taxa de compressão de 80%.

Na figura a seguir (média acumulada de F-Measure) percebe-se que o BLMSumm e o GistSumm apresentam ligeira vantagem.

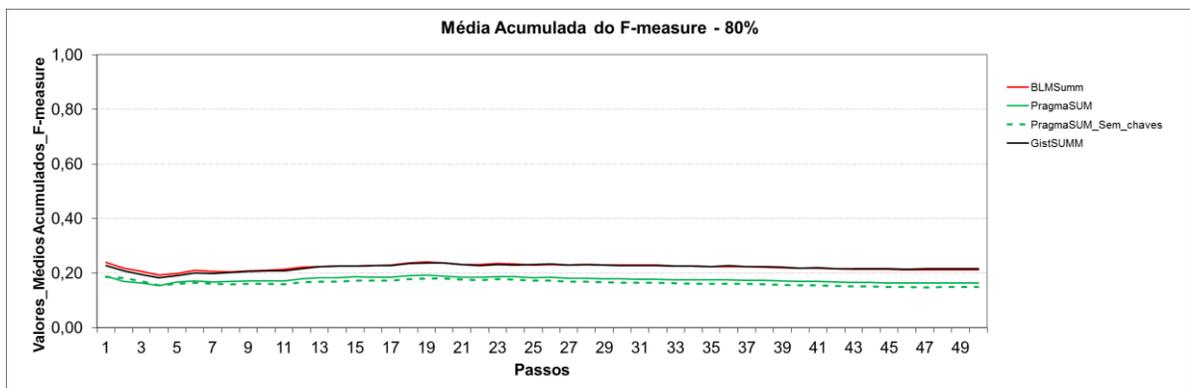


Figura 80. Média acumulada de F-Measure com taxa de compressão de 80% do domínio Engenharia.

A figura a seguir (média acumulada de Precision) mostram que, de uma maneira geral, as taxas de precisão alcançadas foram altas e que o sumarizador que obteve os melhores valores foi o PragmaSUM.

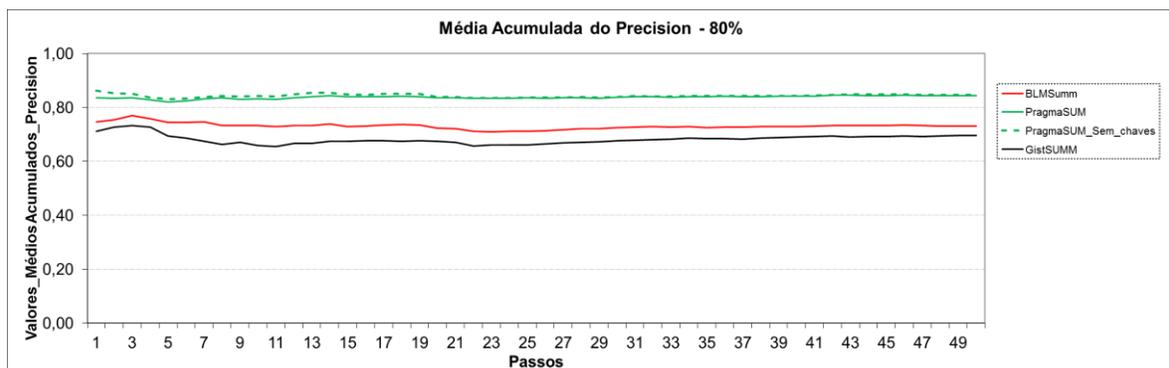


Figura 81. Média acumulada de Precision com taxa de compressão de 80% do domínio Engenharia.

A figura a seguir (média acumulada de Recall) evidenciam que quanto à cobertura os sumarizadores obtiveram valores muito próximos com ligeira vantagem para o BLMSumm e o GistSumm.

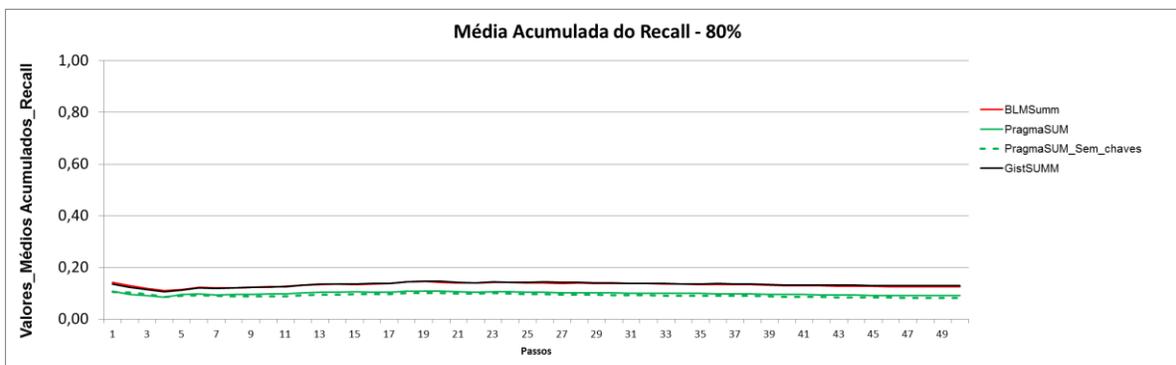


Figura 82. Média acumulada de Recall com taxa de compressão de 80% do domínio Engenharia.

4.4. Taxa de Compressão de 90%

Nas figuras desta seção têm-se os resultados das SA para os 50 textos do domínio Engenharia com taxa de compressão aplicada de 90%.

É possível perceber figura a seguir (média acumulada de F-Measure) que, de forma geral, os sumarizadores obtiveram valores muito próximos com ligeira vantagem para o BLMSumm e o GistSumm.

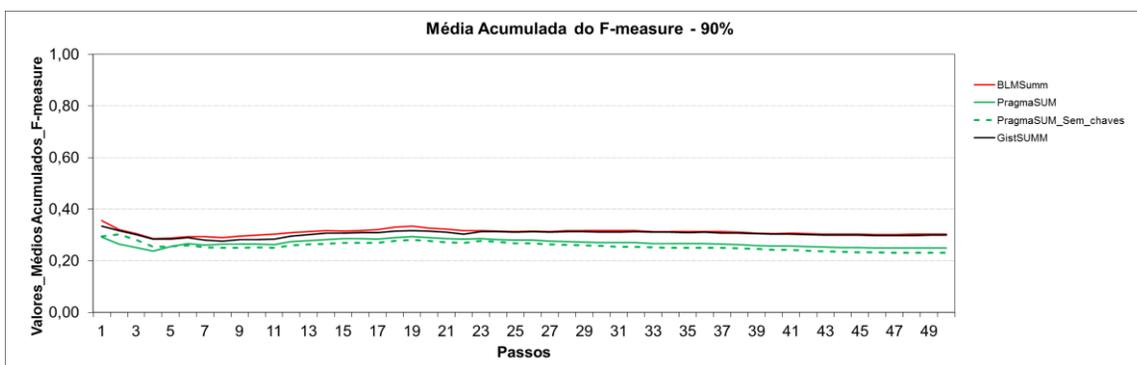


Figura 83. Média acumulada de F-Measure com taxa de compressão de 90% do domínio Engenharia.

É possível perceber na figura a seguir (média acumulada de Precision) que quanto à precisão o PragmaSUM obteve os melhores resultados dentre os sumarizadores analisados.

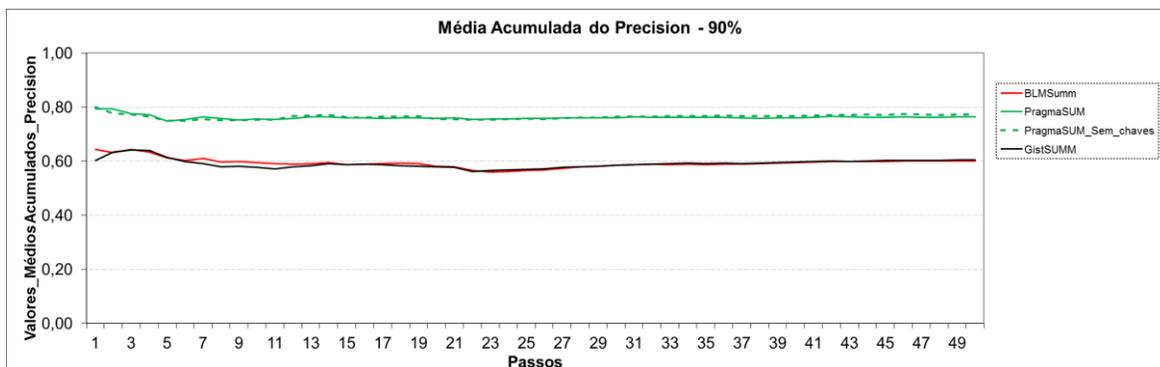


Figura 84. Média acumulada de Precision com taxa de compressão de 90% do domínio Engenharia.

É possível perceber na figura a seguir (média acumulada de Recall) que quanto à cobertura o BLMSumm obteve os melhores resultados junto com o GistSumm, com valores muito próximos entre si.

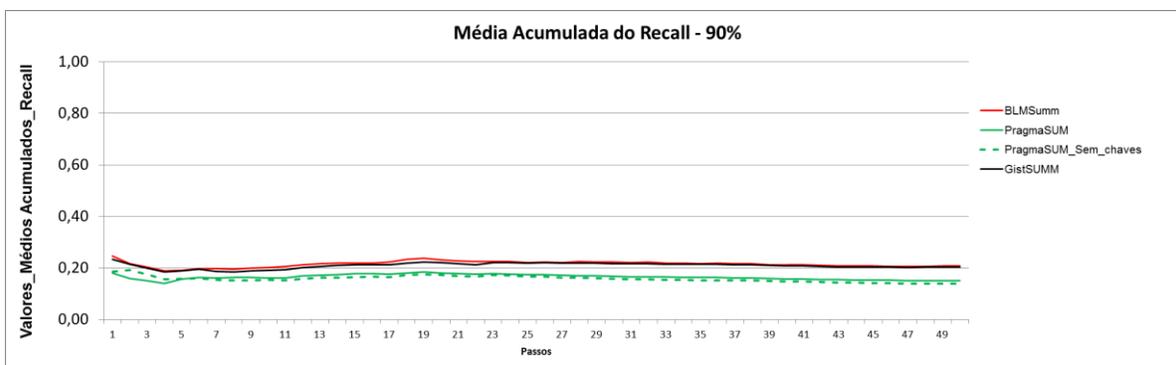


Figura 85. Média acumulada de Recall com taxa de compressão de 90% do domínio Engenharia.

4.5. Comparação dos Resultados

A tabela a seguir mostra uma comparação feita entre as taxas de compressão das médias gerais de média harmônica (F-Measure) alcançadas pelos cinco sumarizadores automáticos para os textos do domínio engenharia. O GistSumm obteve os melhores resultados, nas taxas de 50 e 70%. A Figura 81 reafirma as observações.

Tabela 23. Comparação dos resultados de F-Measure do domínio Engenharia.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,10895	0,19551	0,08617	0,08121

70%	0,16302	0,19854	0,12369	0,11346
80%	0,21243	0,2162	0,16364	0,14849
90%	0,30255	0,29965	0,24917	0,23076

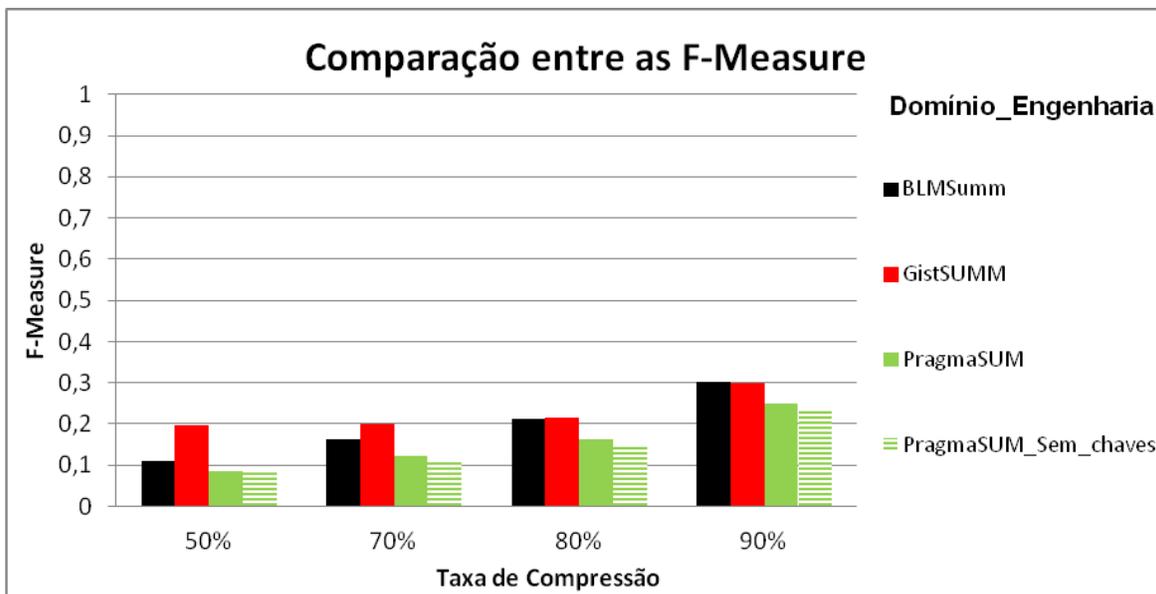


Figura 86. Comparação dos resultados de F-Measure do domínio Engenharia.

A tabela e a figura a seguir trazem uma comparação feita entre as taxas de compressão das médias gerais de Recall (cobertura) que os sumarizadores automáticos alcançaram para os textos do domínio Engenharia. Constata-se que o GistSumm apresentou os melhores resultados para as taxas de 50 e 70%. Nas compressões de 80 e 90% os resultados são mais homogêneos.

Tabela 24. Comparação dos resultados de Recall do domínio Engenharia.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,05844	0,11587	0,04545	0,04274
70%	0,09148	0,11793	0,06702	0,06119
80%	0,12604	0,13034	0,0915	0,08239
90%	0,20728	0,20408	0,15159	0,13942

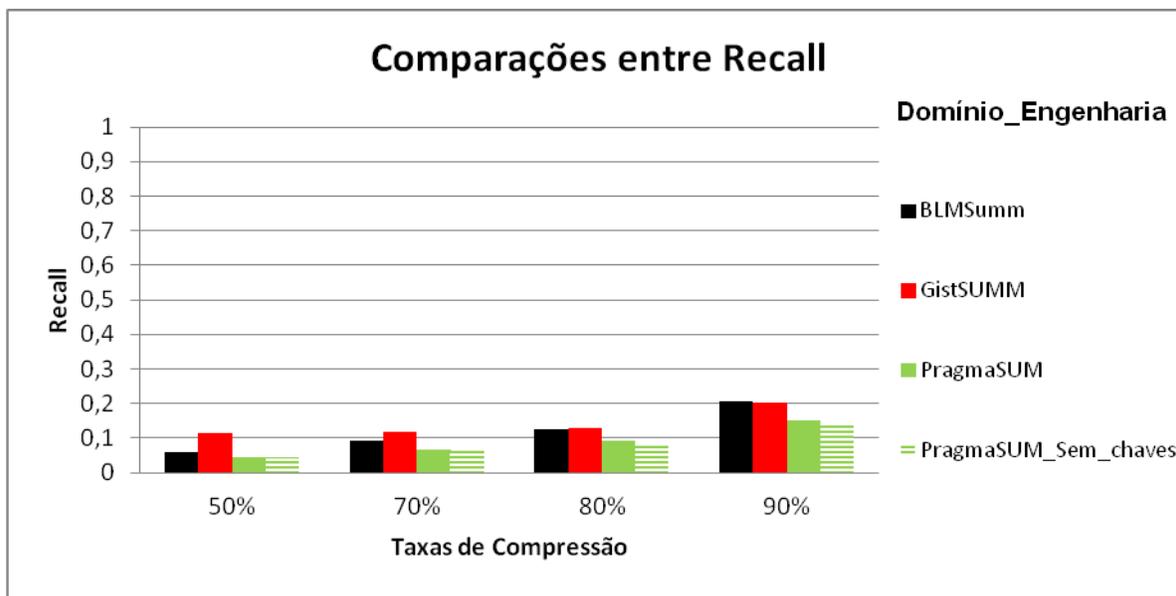


Figura 87. Comparação dos resultados de Recall do domínio Engenharia.

5. Domínio Filosofia

Esta seção descreve os resultados das SA realizadas no domínio Filosofia.

5.1. Taxa de Compressão de 50%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio Filosofia. A taxa de compressão aplicada foi de 50%.

A figura a seguir (média acumulada de F-Measure) apresentam os resultados das SA para cada um dos sumarizadores automáticos. Observa-se que há bastante equilíbrio nos resultados, mas o GistSumm possui uma ligeira vantagem.

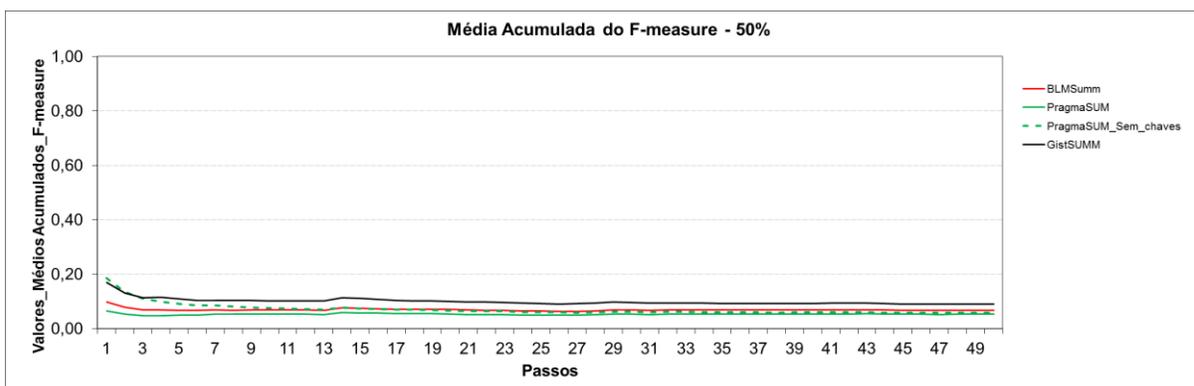


Figura 88. Média acumulada de F-Measure com taxa de compressão de 50% do domínio Filosofia.

Na figura a seguir (média acumulada de Precision) têm-se os resultados para cada um dos sumarizadores automáticos. Quanto à precisão (Precision) o PragmaSUM apresentou os melhores resultados com uma ligeira vantagem sobre sua versão sem chaves.

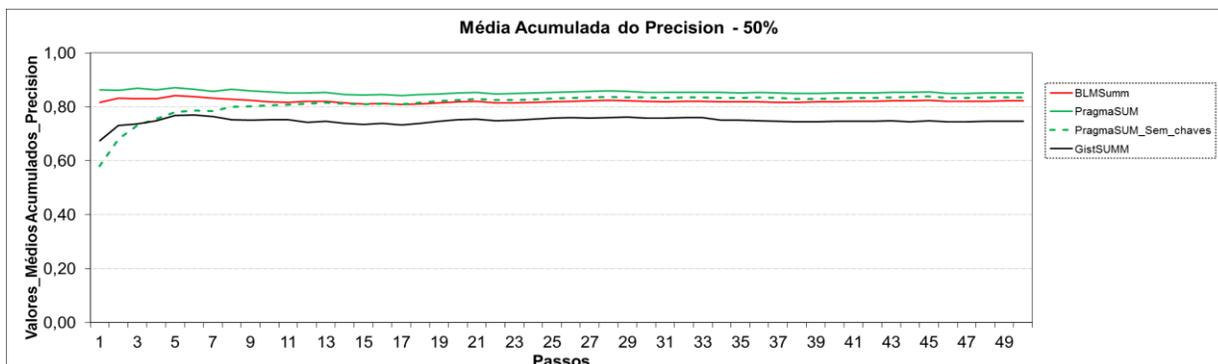


Figura 89. Média acumulada de Precision com taxa de compressão de 50% do domínio Filosofia.

A figura a seguir (média acumulada de Recall) mostram os resultados da cobertura para cada um dos sumarizadores automáticos. É possível perceber que existe uma variação muito pequena dentre resultados aferidos.

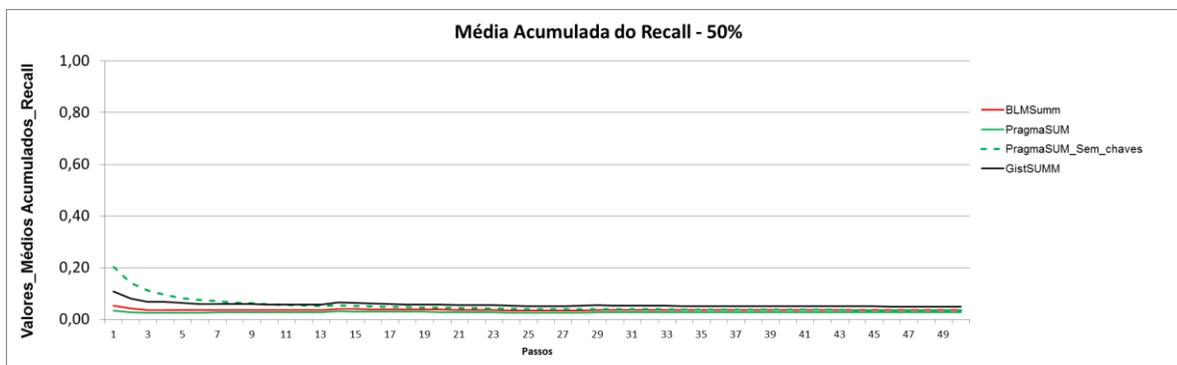


Figura 90. Média acumulada de Recall com taxa de compressão de 50% do domínio Filosofia.

5.2. Taxa de Compressão de 70%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio Filosofia para uma taxa de compressão de 70%.

É possível perceber na figura a seguir (média acumulada de F-Measure) que existe uma variação muito pequena dentre resultados aferidos.

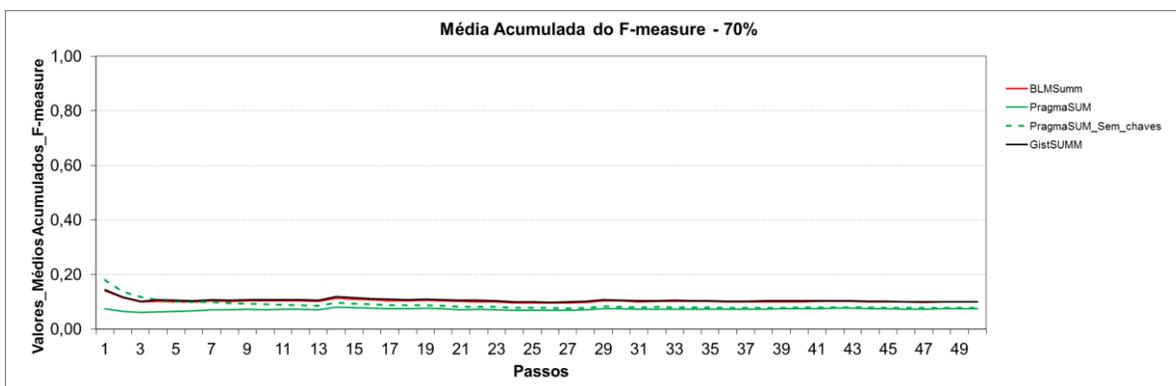


Figura 91. Média acumulada de F-Measure com taxa de compressão de 70% do domínio Filosofia.

Percebe-se na figura a seguir (média acumulada de Precision) que o sumarizador com os valores de precisão mais altos foi o PragmaSUM, com ligeira vantagem sobre sua versão sem chaves.

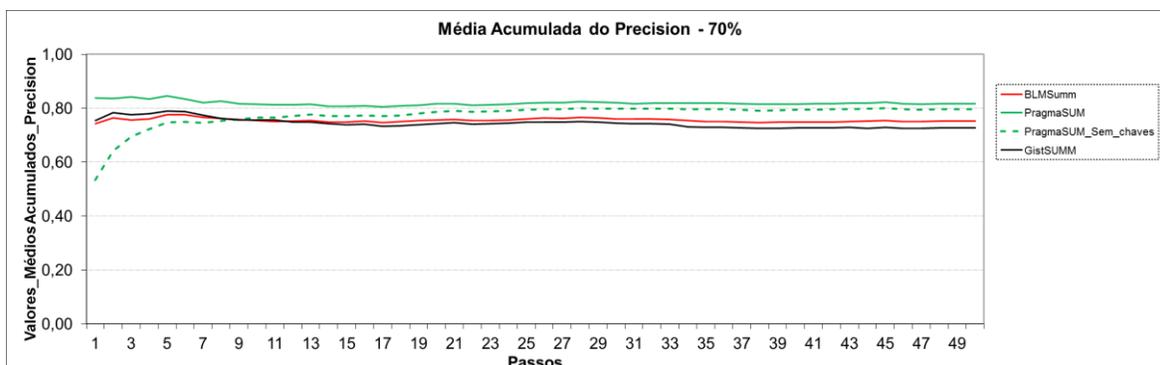


Figura 92. Média acumulada de Precision com taxa de compressão de 70% do domínio Filosofia.

A figura a seguir (média acumulada de Recall) mostram que quanto à cobertura os sumarizadores que existe uma variação muito pequena dentre resultados aferidos.

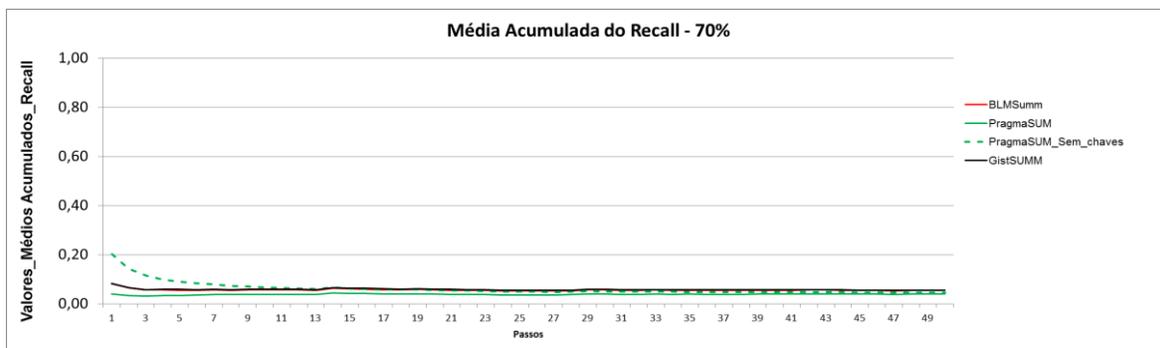


Figura 93. Média acumulada de Recall com taxa de compressão de 70% do domínio Filosofia.

5.3. Taxa de Compressão de 80%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio Filosofia com taxa de compressão de 80%.

Na figura a seguir (média acumulada de F-Measure) percebe-se que o BLMSumm apresenta os resultados mais satisfatórios junto com o GistSumm, ligeiramente melhores do que o PragmaSUM.

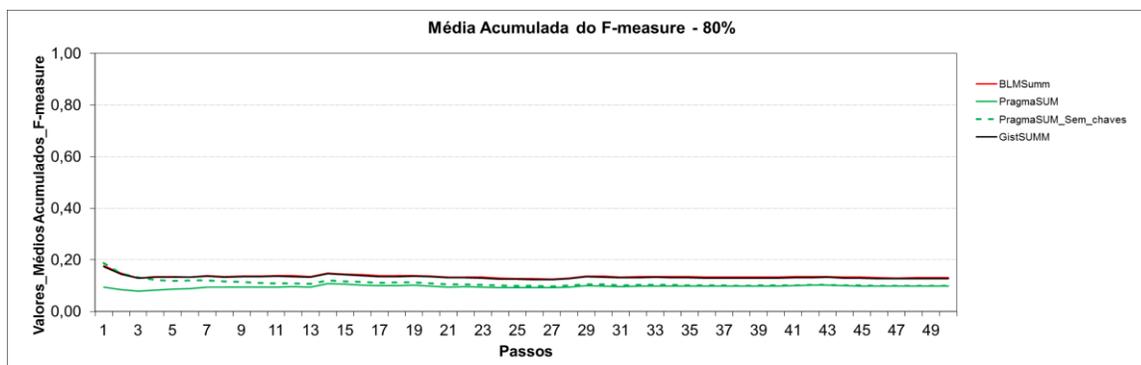


Figura 94. Média acumulada de F-Measure com taxa de compressão de 80% do domínio Filosofia.

A figura a seguir (média acumulada de Precision) mostram que, de uma maneira geral, as taxas de precisão alcançadas foram altas e que o sumarizador que obteve os melhores valores foi o PragmaSUM, seguido por sua versão sem chaves.

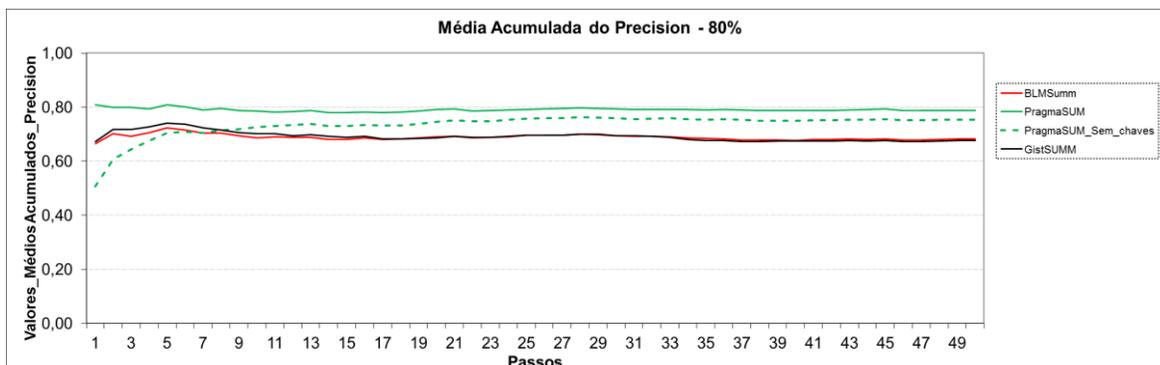


Figura 95. Média acumulada de Precision com taxa de compressão de 80% do domínio Filosofia.

A figura a seguir (média acumulada de Recall) evidenciam que quanto à cobertura os valores apresentam pouca variação entre si.

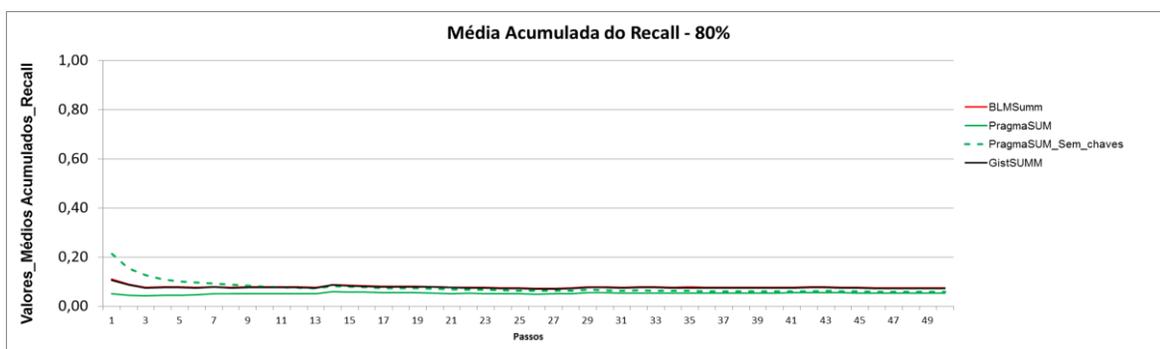


Figura 96. Média acumulada de Recall com taxa de compressão de 80% do domínio Filosofia.

5.4. Taxa de Compressão de 90%

Nas figuras desta seção têm-se os resultados das SA para os 50 textos do domínio Filosofia com taxa de compressão aplicada de 90%.

É possível perceber na figura a seguir (média acumulada de F-Measure) que, de forma geral, o BLMSumm apresenta os resultados mais satisfatórios seguido proximamente pelo GistSumm.

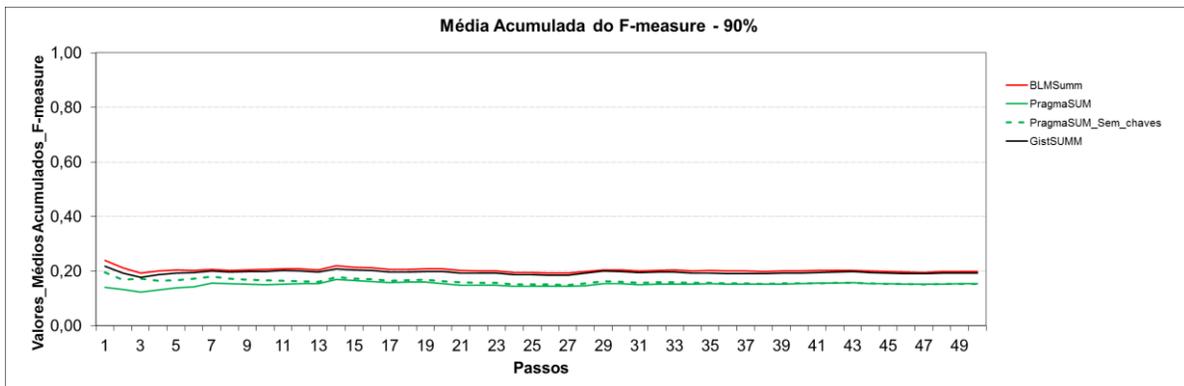


Figura 97. Média acumulada de F-Measure com taxa de compressão de 90% do domínio Filosofia.

É possível perceber na figura a seguir (média acumulada de Precision) que de uma maneira geral, as taxas de precisão alcançadas foram altas e que o sumarizador que obteve os melhores valores foi o PragmaSUM, seguido por sua versão sem chaves.

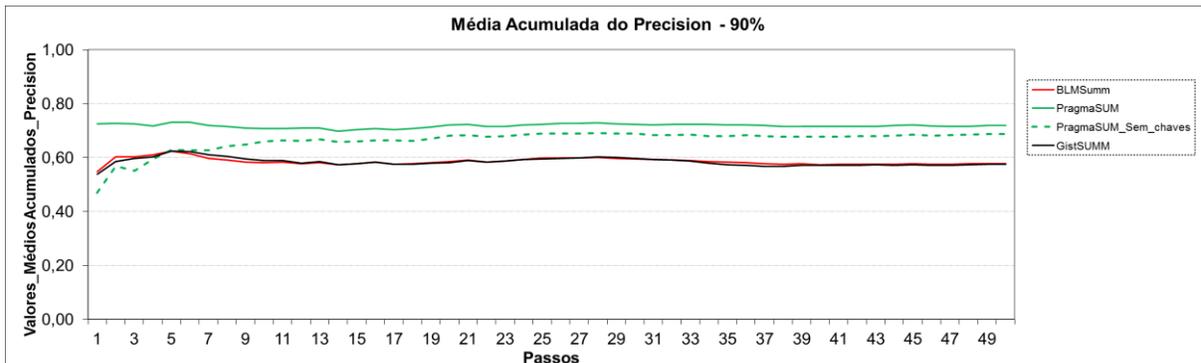


Figura 98. Média acumulada de Precision com taxa de compressão de 90% do domínio Filosofia.

É possível perceber na figura a seguir (média acumulada de Recall) que quanto à cobertura o valores são muito próximos entre si.

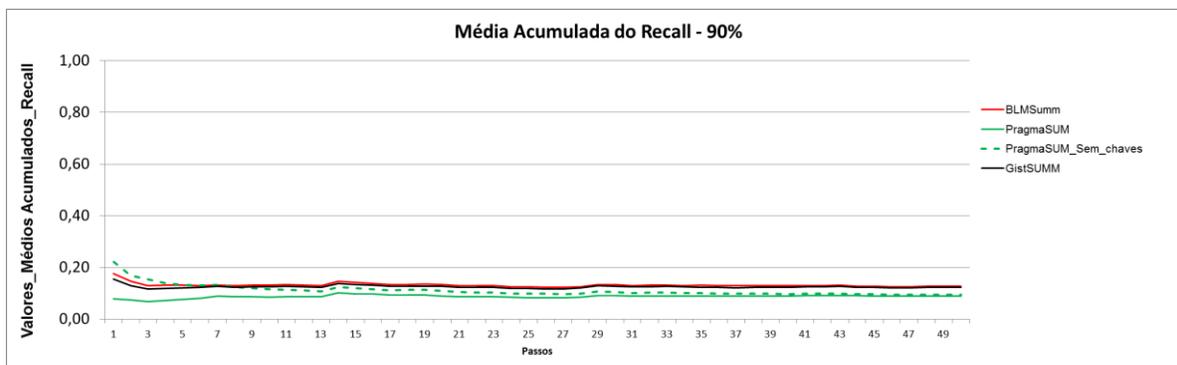


Figura 99. Média acumulada de Recall com taxa de compressão de 90% do domínio Filosofia.

5.5. Comparação dos Resultados

A tabela a seguir mostra uma comparação feita entre as taxas de compressão das médias gerais de média harmônica (F-Measure) alcançadas pelos sumarizadores automáticos para os textos do domínio Filosofia. O BLMSumm obteve os melhores resultados Junto com o GistSumm. Quando a taxa de compressão é de 50% os valores variam pouco entre si. A Figura 100 reafirma as observações.

Tabela 25. Comparação dos resultados de F-Measure do domínio Filosofia.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,06694	0,08954	0,0527	0,05763
70%	0,09914	0,10045	0,07409	0,07737
80%	0,13124	0,12781	0,0988	0,09933
90%	0,19841	0,1931	0,1539	0,15318

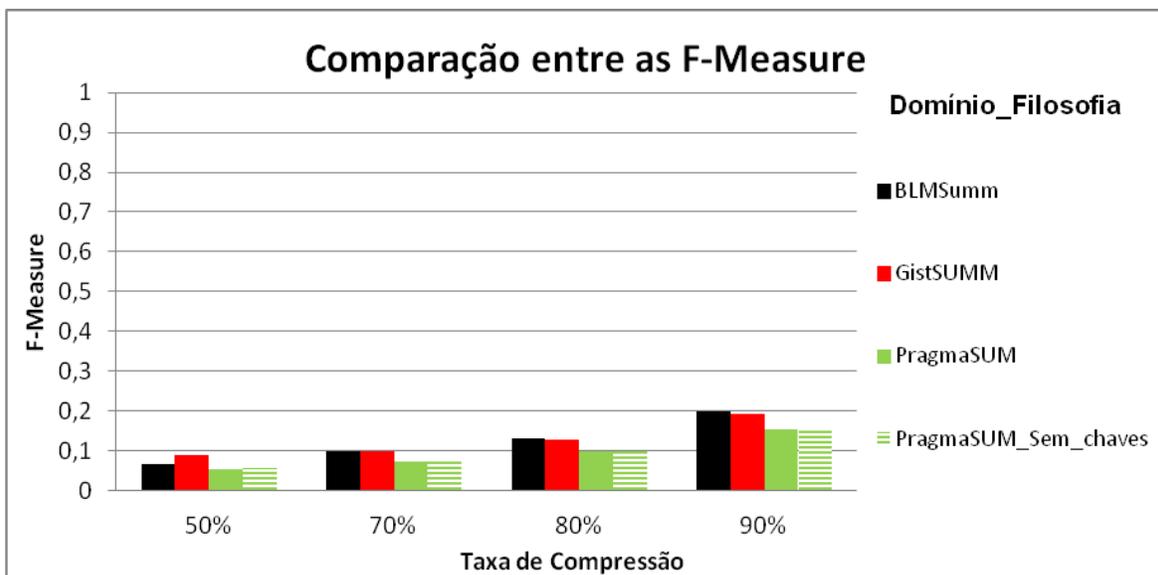


Figura 100. Comparação dos resultados de F-Measure do domínio Filosofia.

A tabela e a figura a seguir trazem uma comparação feita entre as taxas de compressão das médias gerais de Recall (cobertura) que os sumarizadores automáticos alcançaram para os textos do domínio Filosofia. Constata-se que o BLMSumm e o GistSumm apresentaram resultados ligeiramente melhores para as 4 taxas de compressão aplicadas. Os valores apresentados foram muito homogêneos entre si.

Tabela 26. Comparação dos resultados de Recall do domínio Filosofia.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,0354	0,04921	0,0275	0,03428
70%	0,05434	0,05529	0,03941	0,04554
80%	0,07514	0,07301	0,05387	0,05896
90%	0,12668	0,12256	0,08975	0,09495

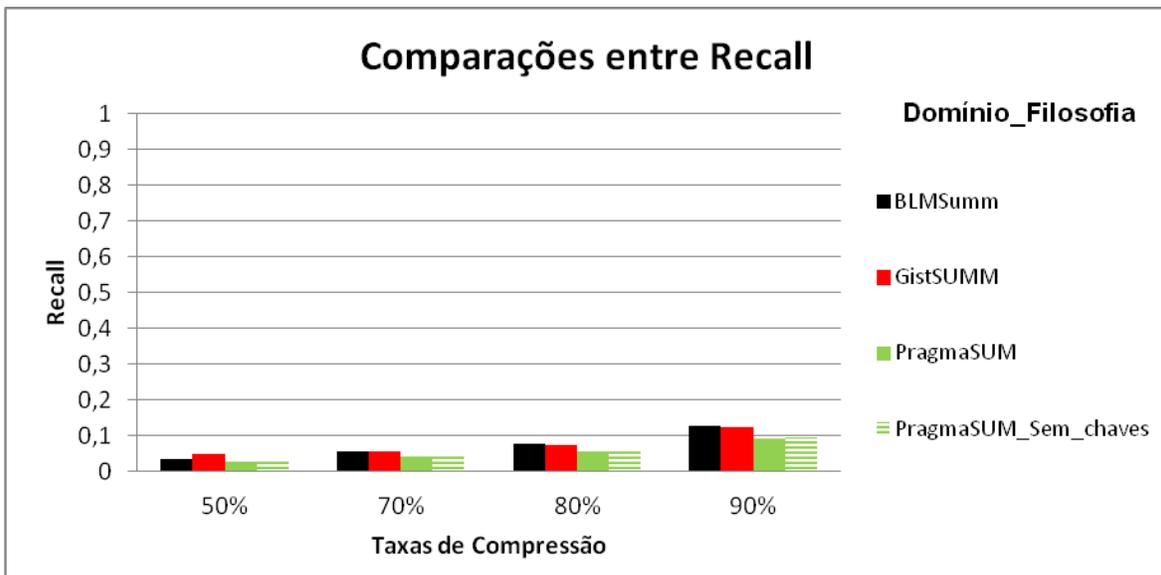


Figura 101. Comparação dos resultados de Recall do domínio Filosofia.

6. Domínio Fisioterapia

Esta seção descreve os resultados das SA realizadas no domínio Fisioterapia.

6.1. Taxa de Compressão de 50%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio Fisioterapia. A taxa de compressão aplicada foi de 50%.

A figura a seguir (média acumulada de F-Measure) apresentam os resultados das SA para cada um dos sumarizadores automáticos. Observa-se que o GistSumm apresenta os melhores resultados.

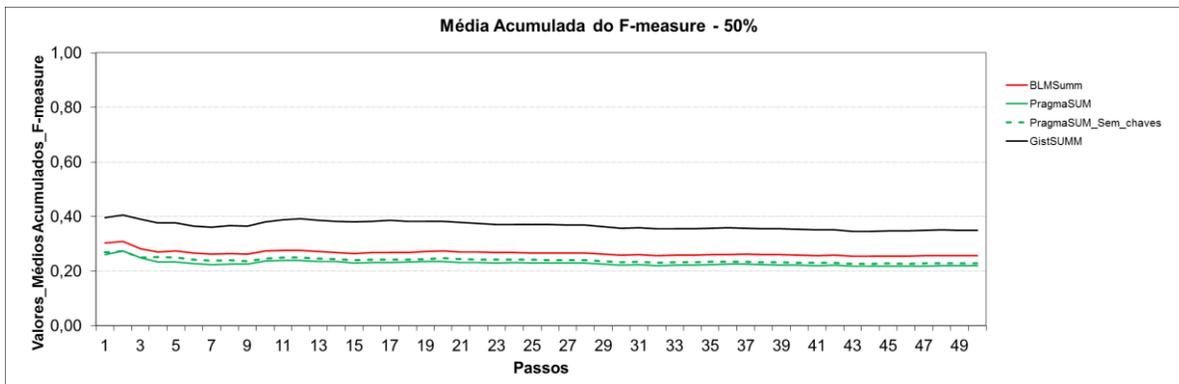


Figura 102. Média acumulada de F-Measure com taxa de compressão de 50% do domínio Fisioterapia.

Na figura a seguir (média acumulada de Precision) têm-se os resultados para cada um dos sumarizadores automáticos. Quanto à precisão (Precision) o PragmaSUM apresentou os melhores resultados com ligeira vantagem sobre o BLMSumm.

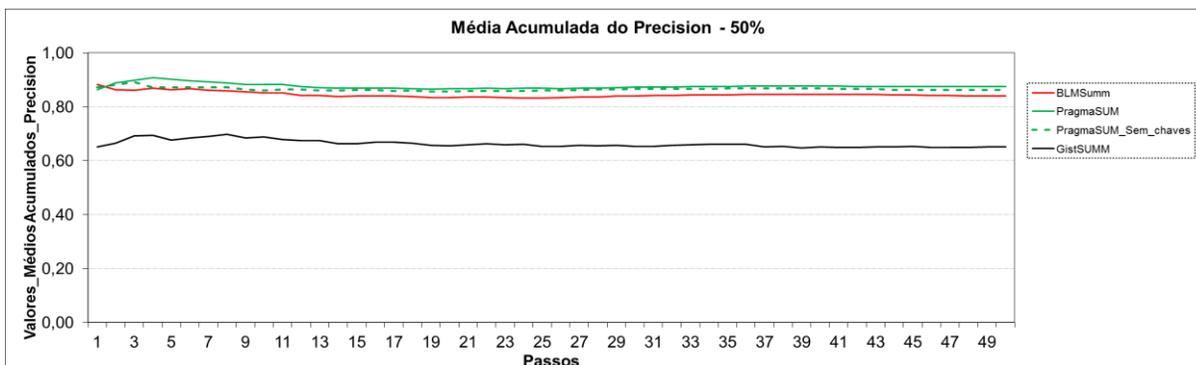


Figura 103. Média acumulada de Precision com taxa de compressão de 50% do domínio Fisioterapia.

A tabela 91 e a figura 91 (média acumulada de Recall) mostram os resultados da cobertura para cada um dos sumarizadores automáticos. É possível perceber que existe uma variação muito pequena dentre resultados aferidos e o GistSumm possui os melhores resultados.

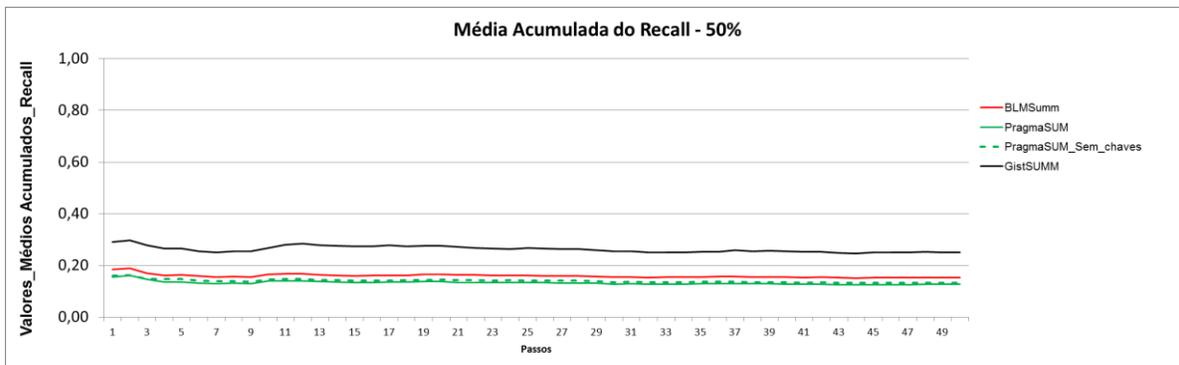


Figura 104. Média acumulada de Recall com taxa de compressão de 50% do domínio Fisioterapia.

6.2. Taxa de Compressão de 70%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio Fisioterapia para uma taxa de compressão de 70%.

É possível perceber na figura a seguir (média acumulada de F-Measure) que o GistSumm apresenta os resultados mais satisfatórios com ligeira vantagem sobre os outros sumarizadores.

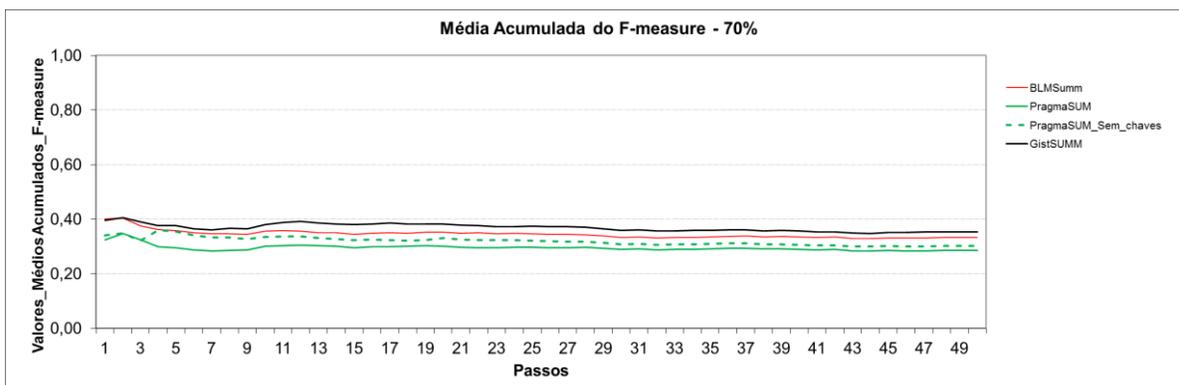


Figura 105. Média acumulada de F-Measure com taxa de compressão de 70% do domínio Fisioterapia.

Percebe-se na figura a seguir (média acumulada de Precision) que o sumarizador com os valores de precisão mais altos foi o PragmaSUM, seguido pelo sua versão sem chaves com ligeira vantagem e o com valores mais baixos foi o GistSumm.

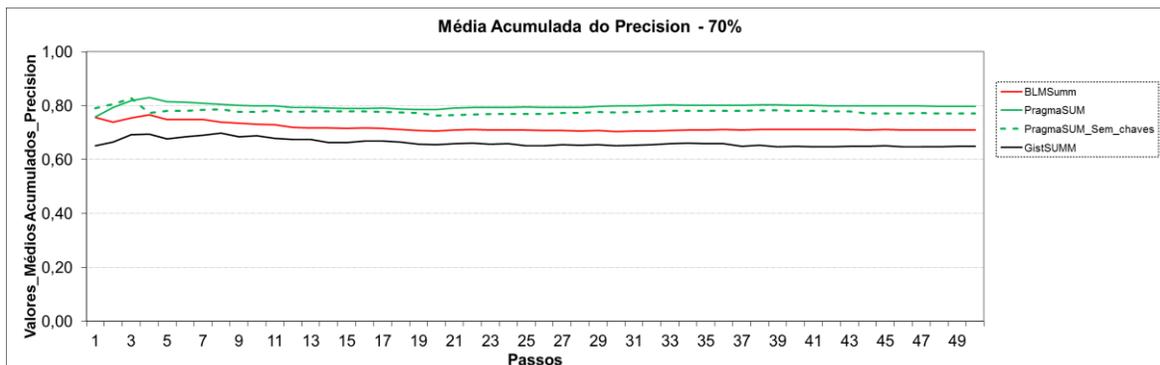


Figura 106. Média acumulada de Precision com taxa de compressão de 70% do domínio Fisioterapia.

A figura a seguir (média acumulada de Recall) mostram que quanto à cobertura os sumarizador que obteve os melhores resultados foi o GistSumm com ligeira vantagem sobre os demais.

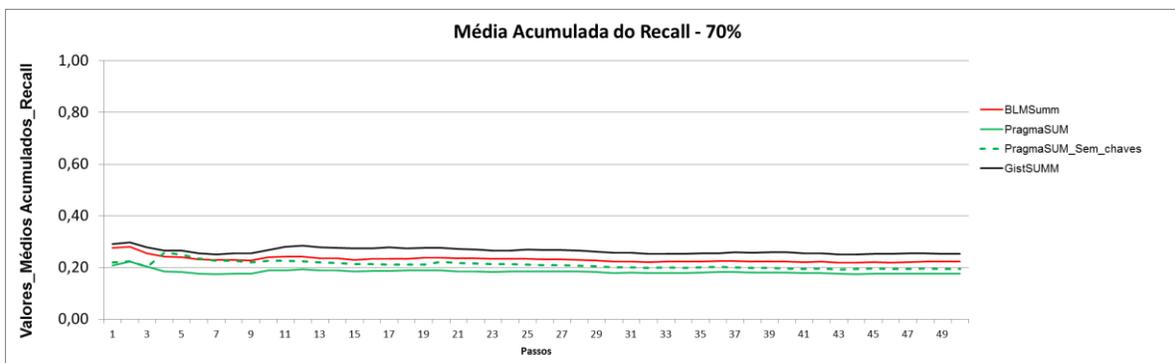


Figura 107. Média acumulada de Recall com taxa de compressão de 70% do domínio Fisioterapia.

6.3. Taxa de Compressão de 80%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio Fisioterapia com taxa de compressão de 80%.

Na figura a seguir (média acumulada de F-Measure) percebe-se que o BLMSumm apresenta os resultados mais satisfatórios junto com o GistSumm.

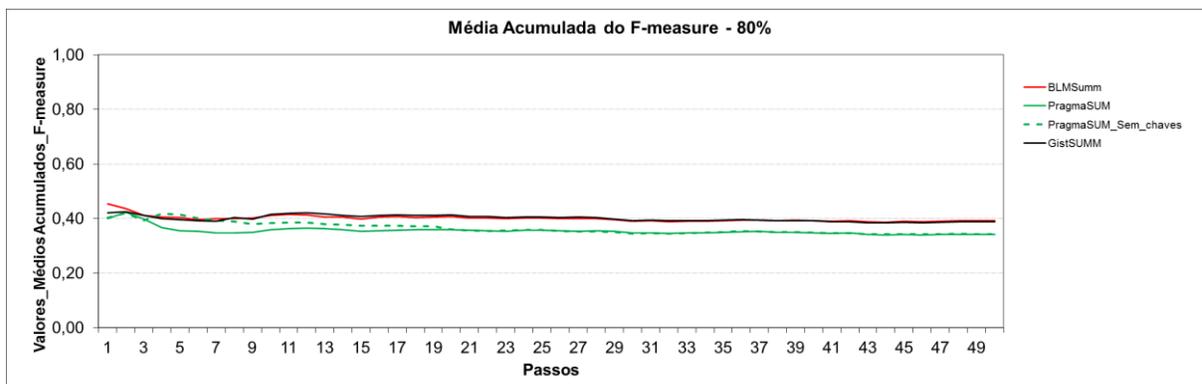


Figura 108. Média acumulada de F-Measure com taxa de compressão de 80% do domínio Fisioterapia.

A figura a seguir (média acumulada de Precision) mostram que, de uma maneira geral, as taxas de precisão alcançadas foram altas e que o sumariizador que obteve os melhores valores foi o PragmaSUM com ligeira vantagem sobre sua versão sem chaves.

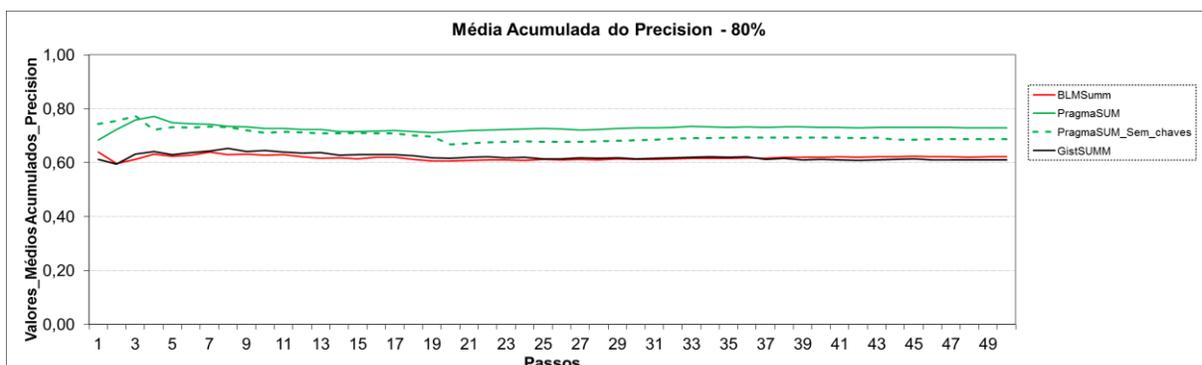


Figura 109. Média acumulada de Precision com taxa de compressão de 80% do domínio Fisioterapia.

A figura a seguir (média acumulada de Recall) evidenciam que quanto à cobertura o BLMSum e o GistSumm obtiveram os melhores resultados.

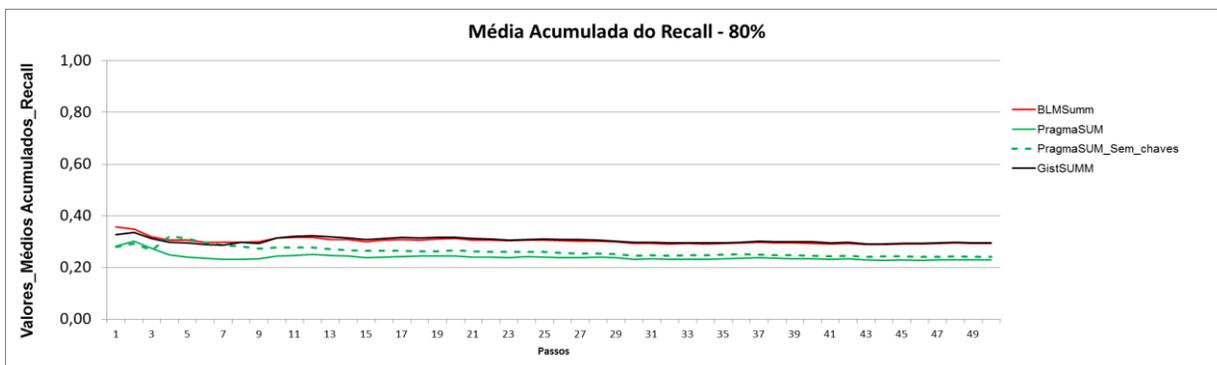


Figura 110. Média acumulada de Recall com taxa de compressão de 80% do domínio Fisioterapia.

6.4. Taxa de Compressão de 90%

Nas figuras desta seção têm-se os resultados das SA para os 50 textos do domínio Fisioterapia com taxa de compressão aplicada de 90%.

É possível perceber na figura a seguir (média acumulada de F-Measure) que, de forma geral, os valores apresentados foram muito próximos entre si.

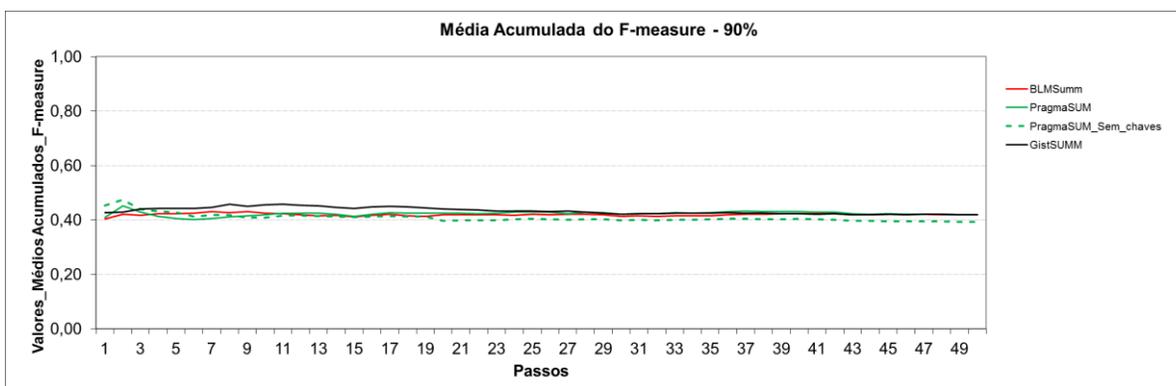


Figura 111. Média acumulada de F-Measure com taxa de compressão de 90% do domínio Fisioterapia.

É possível perceber na figura a seguir (média acumulada de Precision) que quanto à precisão o PragmaSUM obteve os melhores resultados dentre os sumarizadores analisados, com uma boa vantagem sobre sua versão sem chaves. BLMSumm e o GistSumm obtiveram valores próximos entre si.

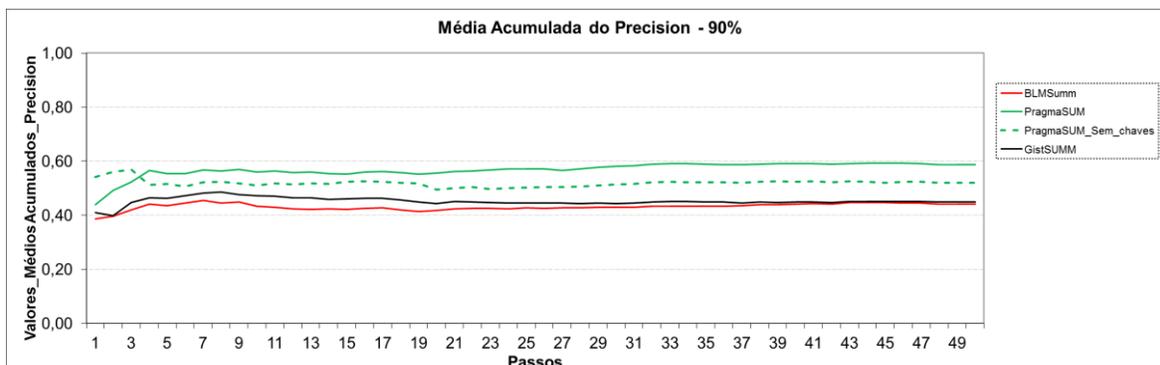


Figura 112. Média acumulada de Precision com taxa de compressão de 90% do domínio Fisioterapia.

É possível perceber na figura a seguir (média acumulada de Recall) que quanto à cobertura o BLMSum e o GistSumm obtiveram ligeira vantagem.

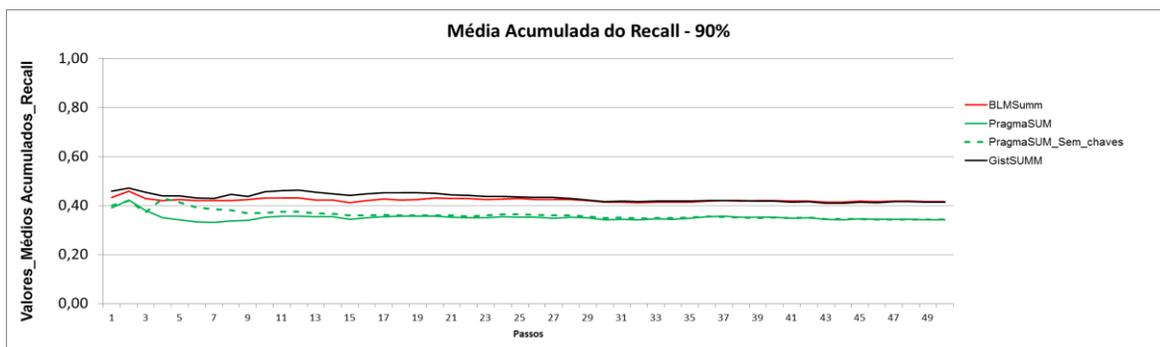


Figura 113. Média acumulada de Recall com taxa de compressão de 90% do domínio Fisioterapia.

6.5. Comparação dos Resultados

A tabela a seguir mostra uma comparação feita entre as taxas de compressão das médias gerais de média harmônica (F-Measure) alcançadas pelos sumarizadores automáticos para os textos do domínio Fisioterapia. O GistSumm obteve os melhores resultados quando a taxa de compressão foi de 50%. Os resultados foram mais homogêneos nas outras compressões ficando muito próximos com a taxa de 90%. A Figura 114 reafirma as observações.

Tabela 27. Comparação dos resultados de F-Measure do domínio Fisioterapia.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,25595	0,34934	0,21926	0,22816
70%	0,33299	0,35215	0,28551	0,30091
80%	0,39141	0,38723	0,34163	0,3436
90%	0,41891	0,42016	0,42006	0,39384

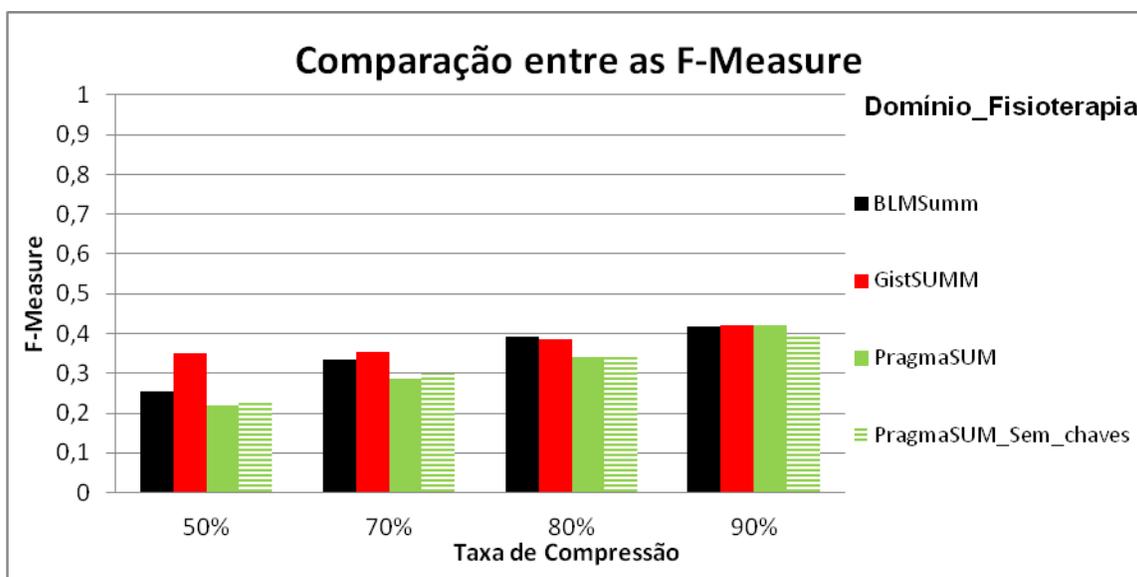


Figura 114. Comparação dos resultados de F-Measure do domínio Fisioterapia.

A tabela e a figura a seguir trazem uma comparação feita entre as taxas de compressão das médias gerais de Recall (cobertura) que os sumarizadores automáticos alcançaram para os textos do domínio Fisioterapia. Constata-se que o GistSumm apresentou os melhores resultados para as taxas de 50 e 70% de compressão aplicadas. O BLMSumm esteve em segundo lugar, exceto para taxa de compressão de 90% onde ele teve ligeira vantagem.

Tabela 28. Comparação dos resultados de Recall do domínio Fisioterapia.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,15279	0,25042	0,12664	0,13288
70%	0,2216	0,25294	0,17702	0,19454
80%	0,29327	0,29486	0,22862	0,24186
90%	0,41652	0,41468	0,34268	0,34188

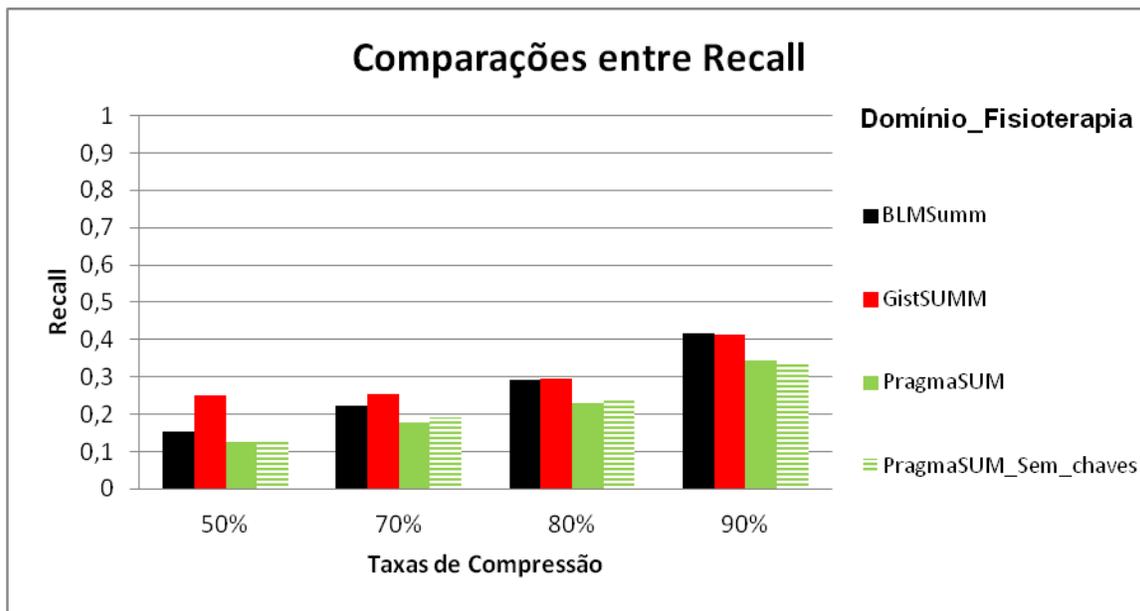


Figura 115. Comparação dos resultados de Recall do domínio Fisioterapia.

7. Domínio Geografia

Esta seção descreve os resultados das SA realizadas no domínio Geografia.

7.1. Taxa de Compressão de 50%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio Geografia. A taxa de compressão aplicada foi de 50%.

A figura a seguir (média acumulada de F-Measure) apresentam os resultados das SA para cada um dos sumarizadores automáticos. Observa-se que há bastante equilíbrio nos resultados, mas o GistSumm possui vantagem.

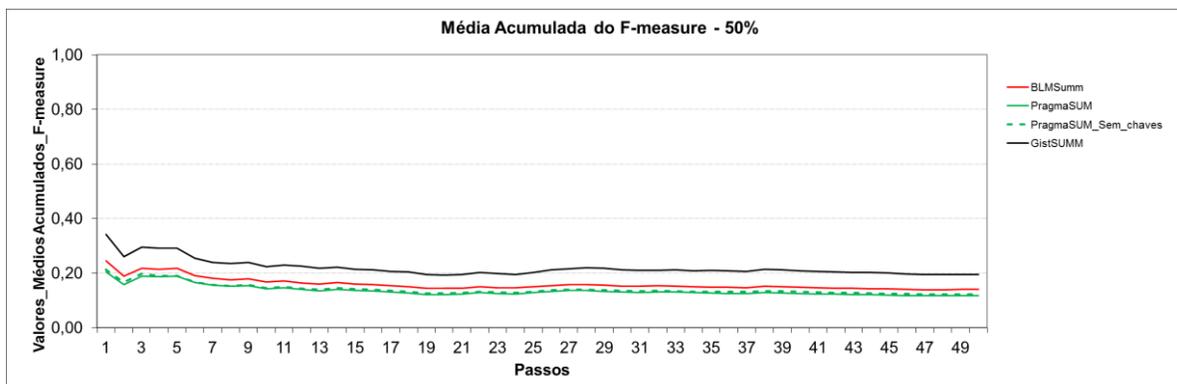


Figura 116. Média acumulada de F-Measure com taxa de compressão de 50% do domínio Geografia.

Na figura a seguir (média acumulada de Precision) têm-se os resultados para cada um dos sumarizadores automáticos. Quanto à precisão (Precision) o PragmaSUM apresentou os melhores resultados.

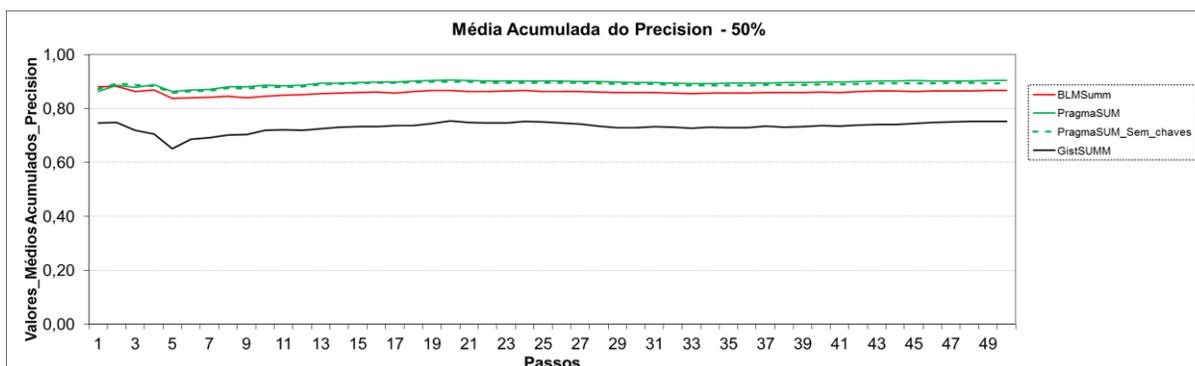


Figura 117. Média acumulada de Precision com taxa de compressão de 50% do domínio Geografia.

A figura a seguir (média acumulada de Recall) mostram os resultados da cobertura para cada um dos sumarizadores automáticos. É possível perceber que existe uma variação muito pequena dentre resultados aferidos e o GistSumm possui ligeira vantagem.

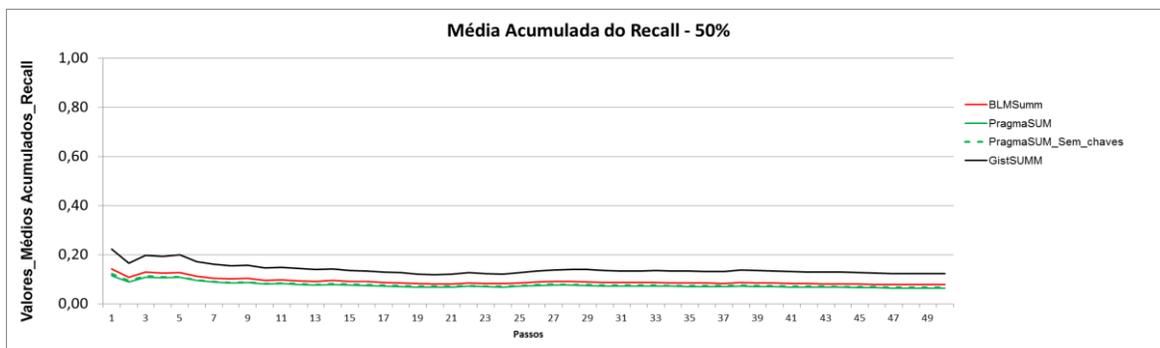


Figura 118. Média acumulada de Recall com taxa de compressão de 50% do domínio Geografia.

7.2. Taxa de Compressão de 70%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio Geografia para uma taxa de compressão de 70%.

É possível perceber na figura a seguir (média acumulada de F-Measure) que o há pouca variação entre os resultados.

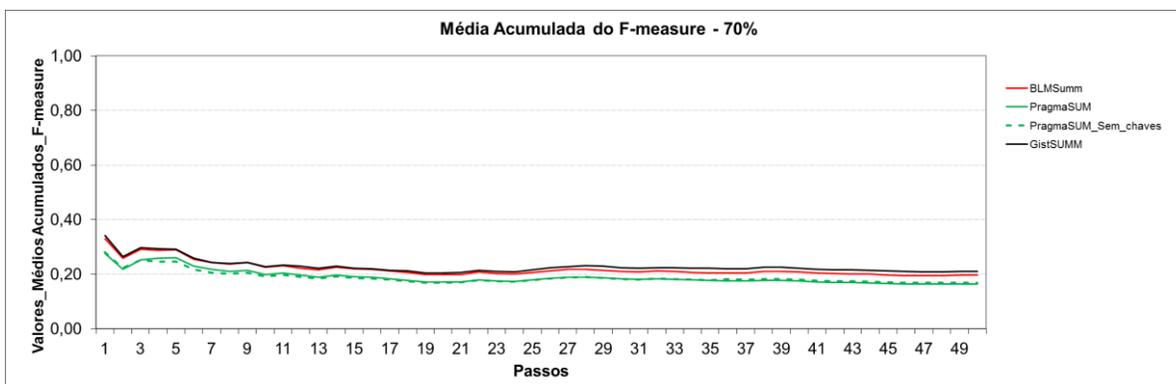


Figura 119. Média acumulada de F-Measure com taxa de compressão de 70% do domínio Geografia.

Percebe-se na figura a seguir (média acumulada de Precision) que o sumarizador com os valores de precisão mais altos foi o PragmaSUM, seguido pela sua versão sem chaves e o com valores mais baixos foi o GistSumm.

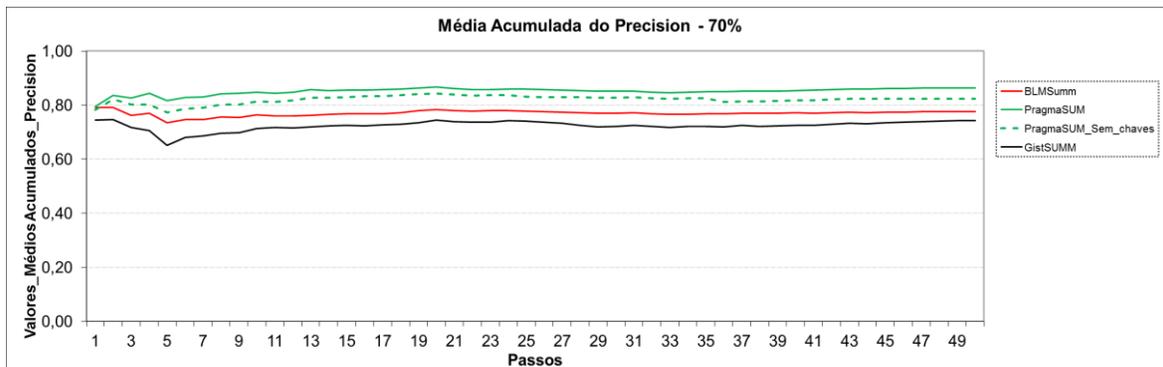


Figura 120. Média acumulada de Precision com taxa de compressão de 70% do domínio Geografia.

A figura a seguir (média acumulada de Recall) mostram que quanto à cobertura os sumarizadores obtiveram pouca variação entre si em seus resultados.

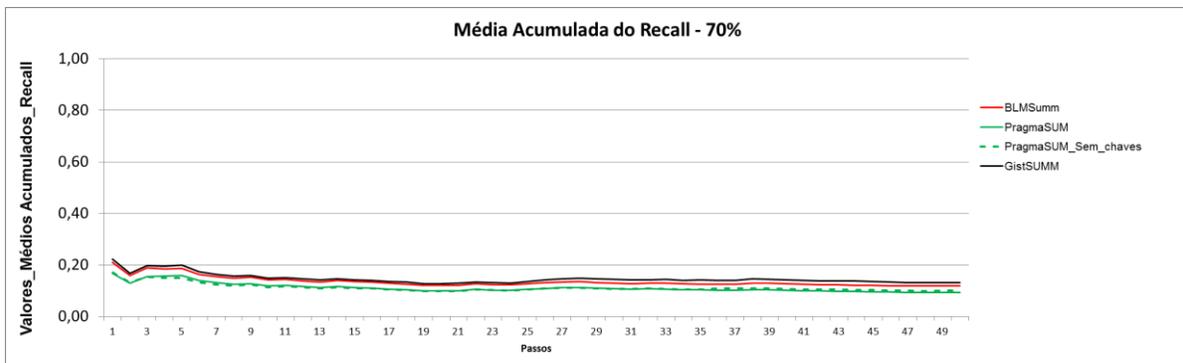


Figura 121. Média acumulada de Recall com taxa de compressão de 70% do domínio Geografia.

7.3. Taxa de Compressão de 80%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio Geografia com taxa de compressão de 80%.

Na figura a seguir (média acumulada de F-Measure) percebe-se que o BLMSumm apresenta os resultados mais satisfatórios junto com o GistSumm com ligeira vantagem sobre o PragmaSUM.

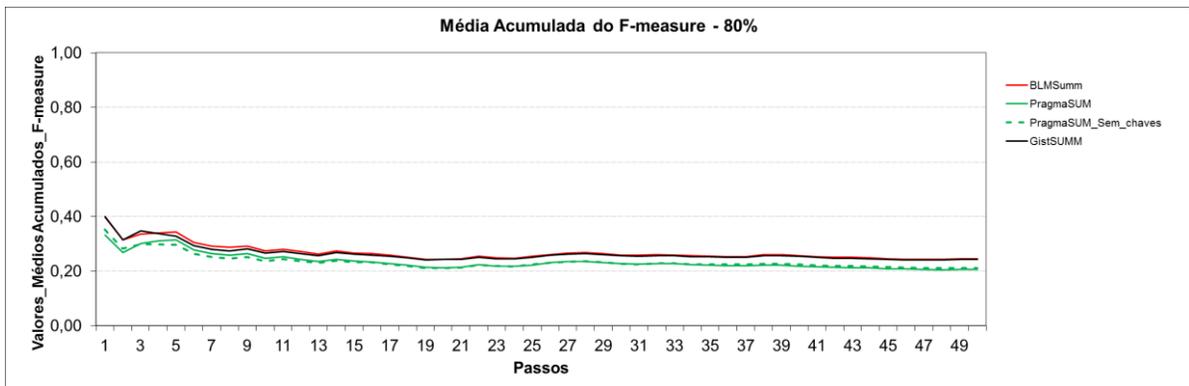


Figura 122. Média acumulada de F-Measure com taxa de compressão de 80% do domínio Geografia.

A figura a seguir (média acumulada de Precision) mostram que, de uma maneira geral, as taxas de precisão alcançadas foram altas e que o sumariador que obteve os melhores valores foi o PragmaSUM com ligeira vantagem sobre sua versão sem chaves.

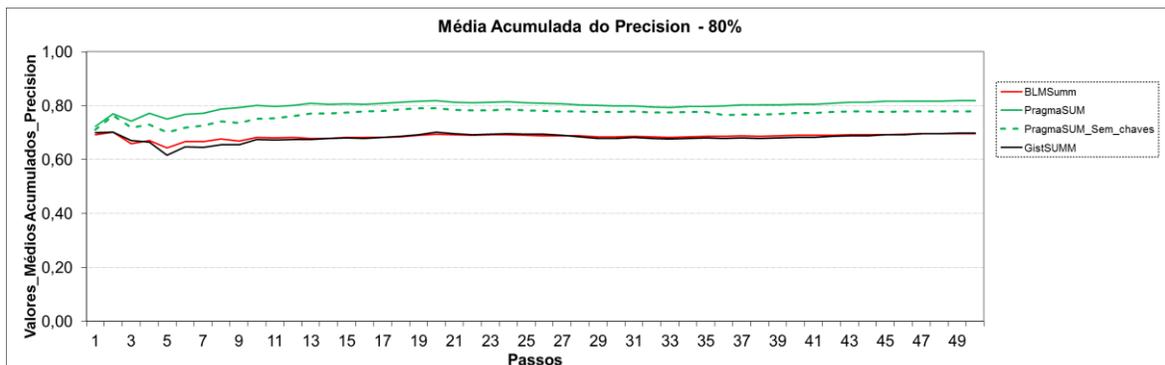


Figura 123. Média acumulada de Precision com taxa de compressão de 80% do domínio Geografia.

A figura a seguir (média acumulada de Recall) evidenciam que quanto à cobertura que os resultados variaram pouco entre si.

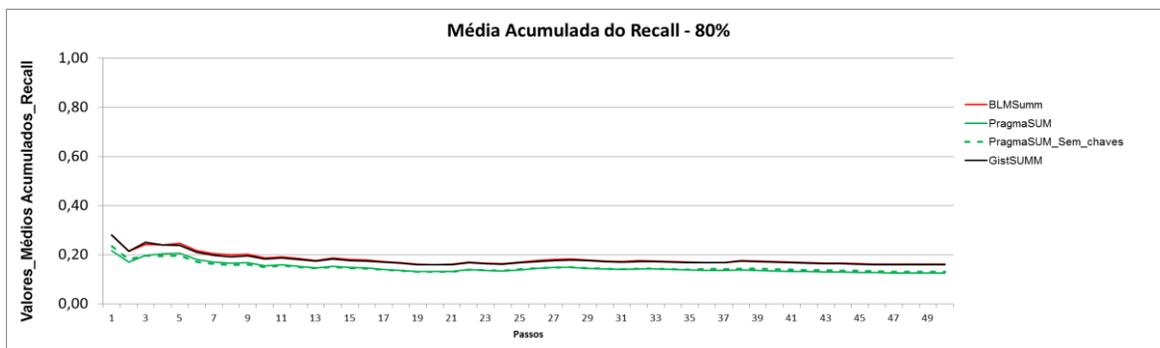


Figura 124. Média acumulada de Recall com taxa de compressão de 80% do domínio Geografia.

7.4. Taxa de Compressão de 90%

Nas figuras desta seção têm-se os resultados das SA para os 50 textos do domínio Geografia com taxa de compressão aplicada de 90%.

É possível perceber na figura a seguir (média acumulada de F-Measure) que, de forma geral, os resultados são muito próximos entre si.

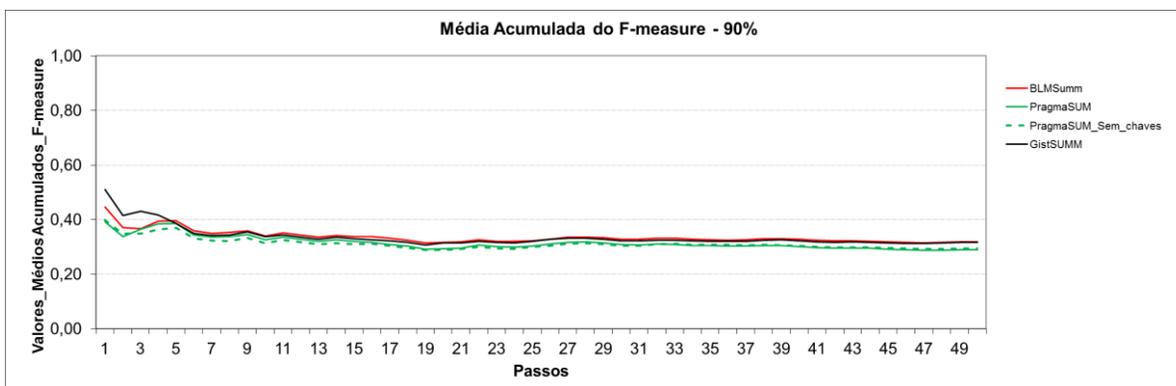


Figura 125. Média acumulada de F-Measure com taxa de compressão de 90% do domínio Geografia.

É possível perceber na figura a seguir (média acumulada de Precision) que quanto à precisão as taxas de precisão alcançadas foram altas e que o sumariador que obteve os melhores valores foi o PragmaSUM com seguido sobre sua versão sem chaves.

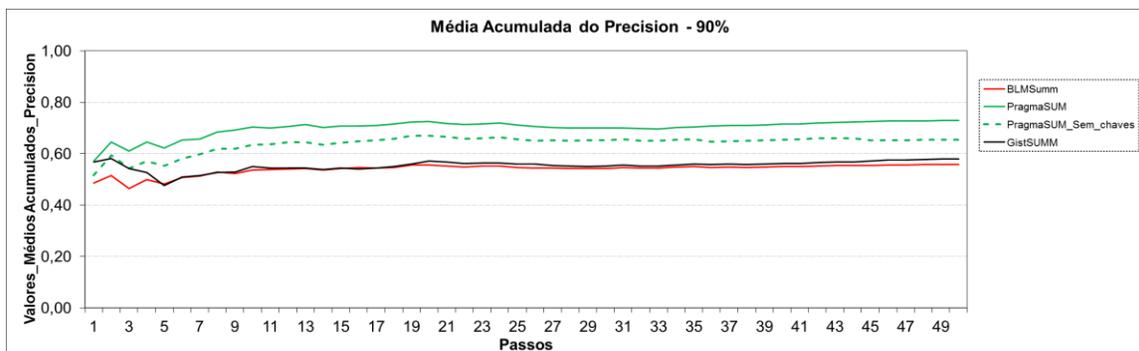


Figura 126. Média acumulada de Precision com taxa de compressão de 90% do domínio Geografia.

É possível perceber na figura a seguir (média acumulada de Recall) que quanto à cobertura o BLMSumm obteve os melhores resultados junto com o GistSumm, mas com pouca variação entre todos os resultados.

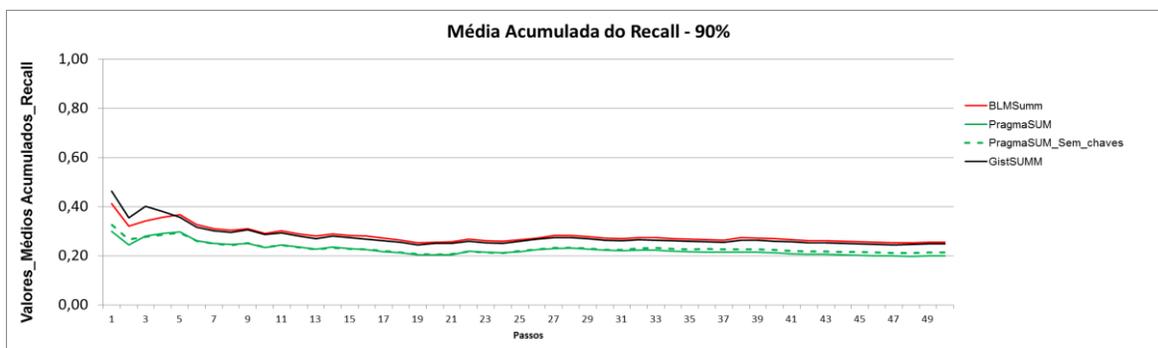


Figura 127. Média acumulada de Recall com taxa de compressão de 90% do domínio Geografia.

7.5. Comparação dos Resultados

A tabela a seguir mostra uma comparação feita entre as taxas de compressão das médias gerais de média harmônica (F-Measure) alcançadas pelos sumarizadores automáticos para os textos do domínio Geografia. O GistSumm obteve os melhores resultados, quando a taxa de compressão foi de 50 e 70%. O BLMSumm tem valores muito próximos do GistSumm com as taxas de 80 e 90%. Os valores variam pouco entre si, exceto com 50%, onde o GistSumm obteve boa vantagem. A Figura 128 reafirma as observações.

Tabela 29. Comparação dos resultados de F-Measure do domínio Geografia.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,13934	0,19486	0,11686	0,12211
70%	0,19488	0,20866	0,16263	0,16773
80%	0,24345	0,24161	0,20539	0,21104
90%	0,31793	0,3162	0,2888	0,29406

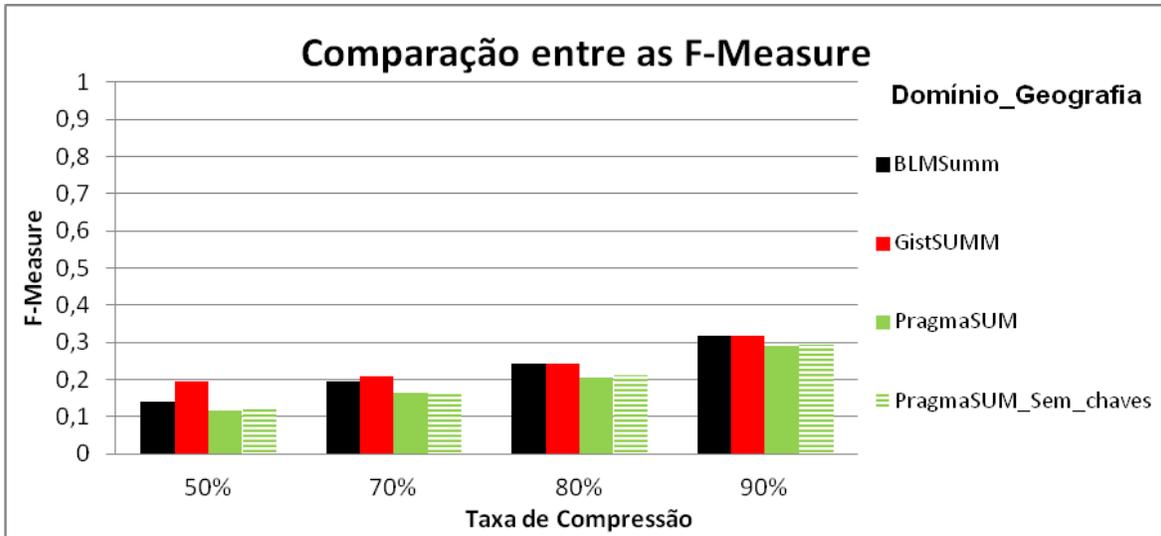


Figura 128. Comparação dos resultados de F-Measure do domínio Geografia.

A tabela e figura a seguir trazem uma comparação feita entre as taxas de compressão das médias gerais de Recall (cobertura) que os sumarizadores automáticos alcançaram para os textos do domínio Geografia. Os valores variam pouco entre si, exceto com 50%, onde o GistSumm obteve ligeira vantagem.

Tabela 30. Comparação dos resultados de Recall do domínio Geografia.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,07845	0,12299	0,06435	0,06745
70%	0,11818	0,13199	0,09391	0,0993
80%	0,16043	0,15872	0,12484	0,13098
90%	0,25317	0,24677	0,19827	0,21214

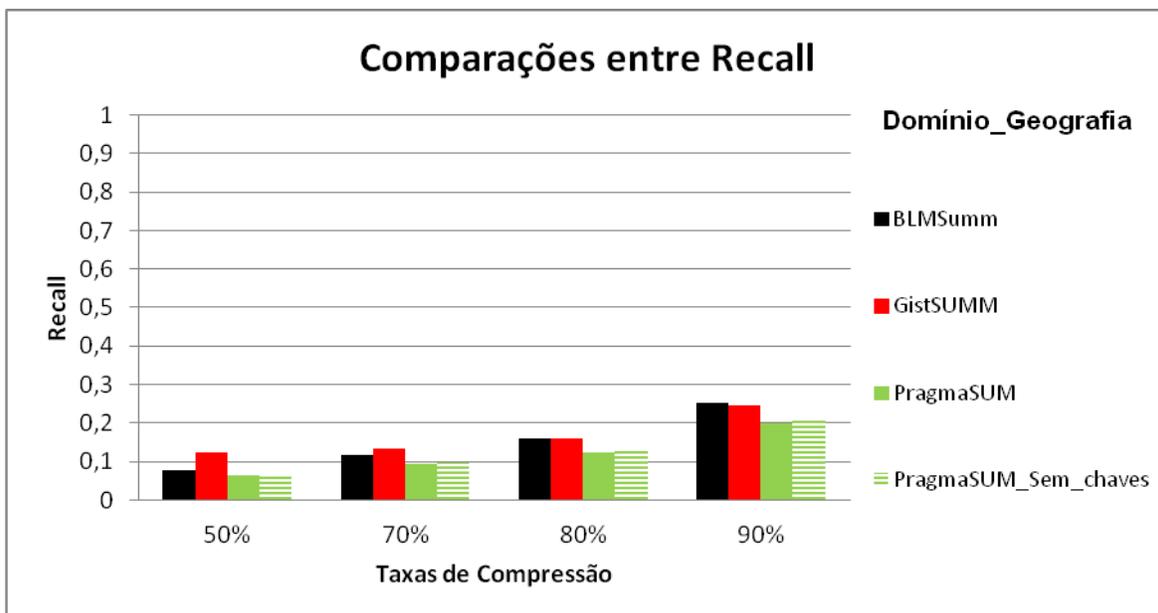


Figura 129. Comparação dos resultados de Recall do domínio Geografia.

8. Domínio História

Esta seção descreve os resultados das SA realizadas no domínio História.

8.1. Taxa de Compressão de 50%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio História. A taxa de compressão aplicada foi de 50%.

A figura a seguir (média acumulada de F-Measure) apresentam os resultados das SA para cada um dos sumarizadores automáticos. Observa-se que há bastante equilíbrio nos resultados, com pouca variação e valores muito baixos e ligeira vantagem para o GistSumm.

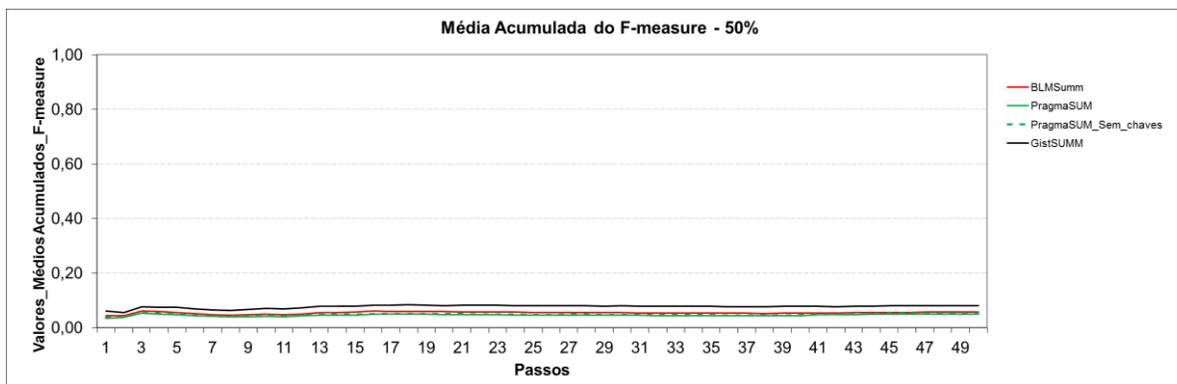


Figura 130. Média acumulada de F-Measure com taxa de compressão de 50% do domínio História.

Na figura a seguir (média acumulada de Precision) têm-se os resultados para cada um dos sumarizadores automáticos. Quanto à precisão (Precision) o PragmaSUM apresentou os melhores resultados.

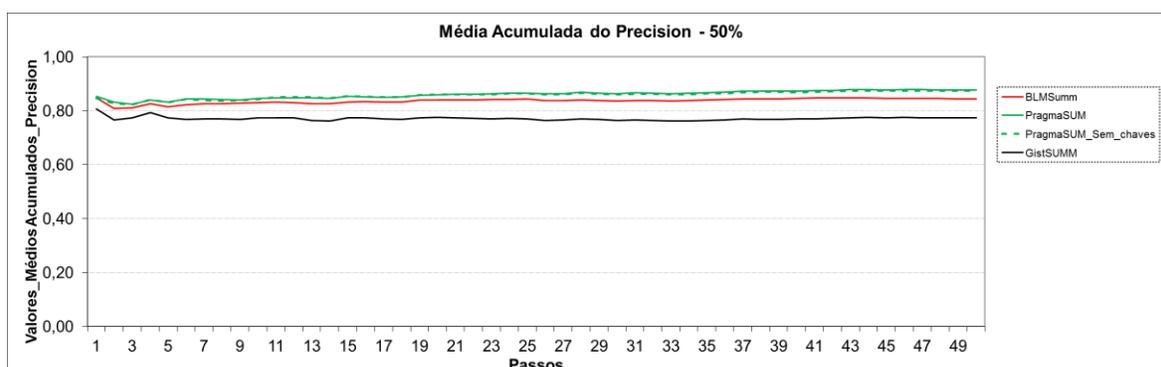


Figura 131. Média acumulada de Precision com taxa de compressão de 50% do domínio História.

A figura a seguir (média acumulada de Recall) mostram os resultados da cobertura para cada um dos sumarizadores automáticos. Observa-se que há bastante equilíbrio nos resultados, com pouca variação e valores muito baixos.

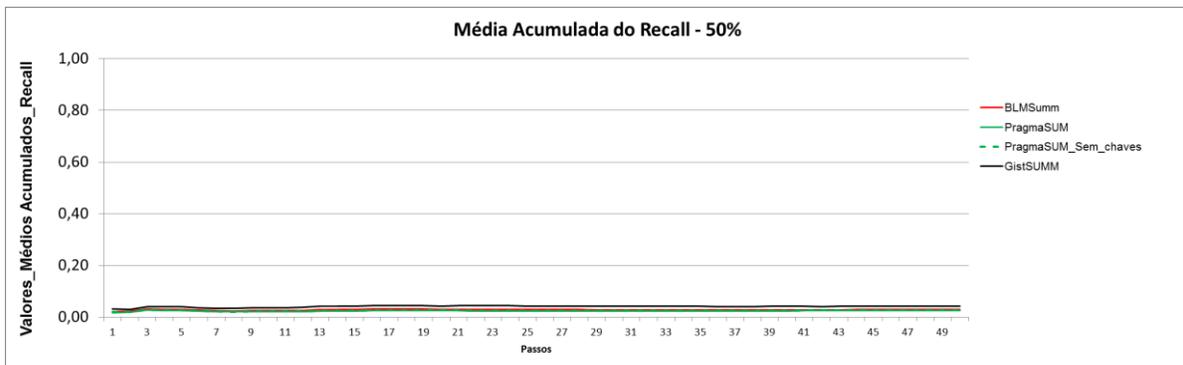


Figura 132. Média acumulada de Recall com taxa de compressão de 50% do domínio História.

8.2. Taxa de Compressão de 70%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio História para uma taxa de compressão de 70%.

É possível perceber na figura a seguir (média acumulada de F-Measure) que há bastante equilíbrio nos resultados, com pouca variação e valores muito baixos.

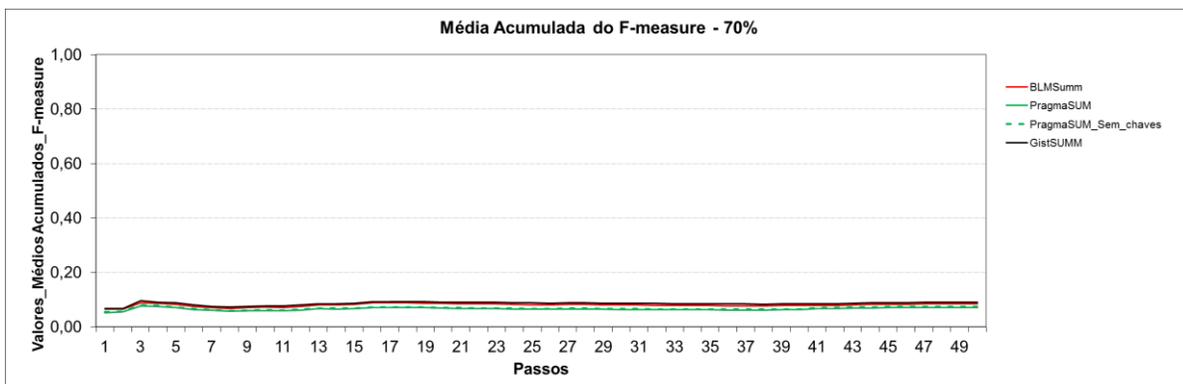


Figura 133. Média acumulada de F-Measure com taxa de compressão de 70% do domínio História.

Percebe-se na figura a seguir (média acumulada de Precision) que o sumarizador com os valores de precisão mais altos foi o PragmaSUM e o com valores mais baixos foi o BLMSumm junto ao GistSumm.

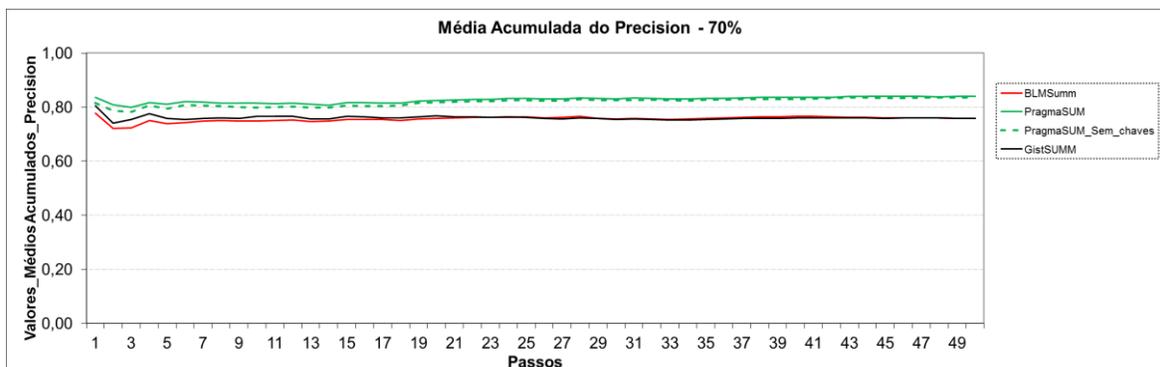


Figura 134. Média acumulada de Precision com taxa de compressão de 70% do domínio História.

A figura a seguir (média acumulada de Recall) mostram que há bastante equilíbrio nos resultados, com pouca variação e valores muito baixos.

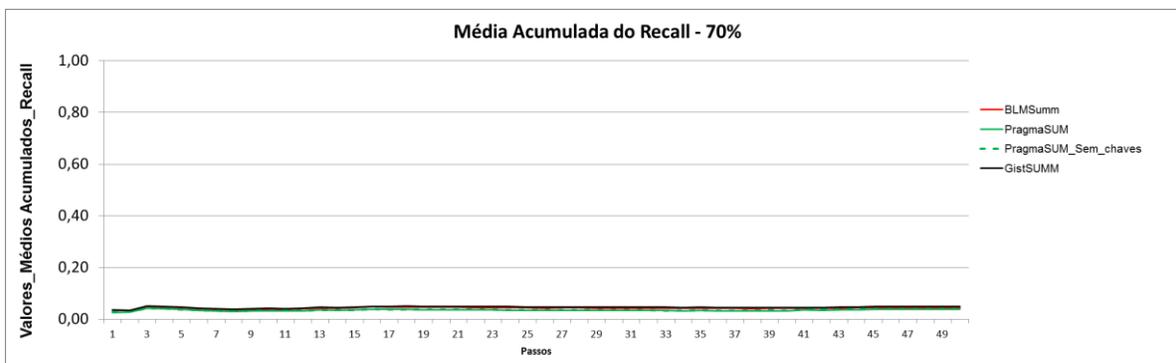


Figura 135. Média acumulada de Recall com taxa de compressão de 70% do domínio História.

8.3. Taxa de Compressão de 80%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio História com taxa de compressão de 80%.

Na figura a seguir (média acumulada de F-Measure) percebe-se que há bastante equilíbrio nos resultados, com pouca variação e valores muito baixos.

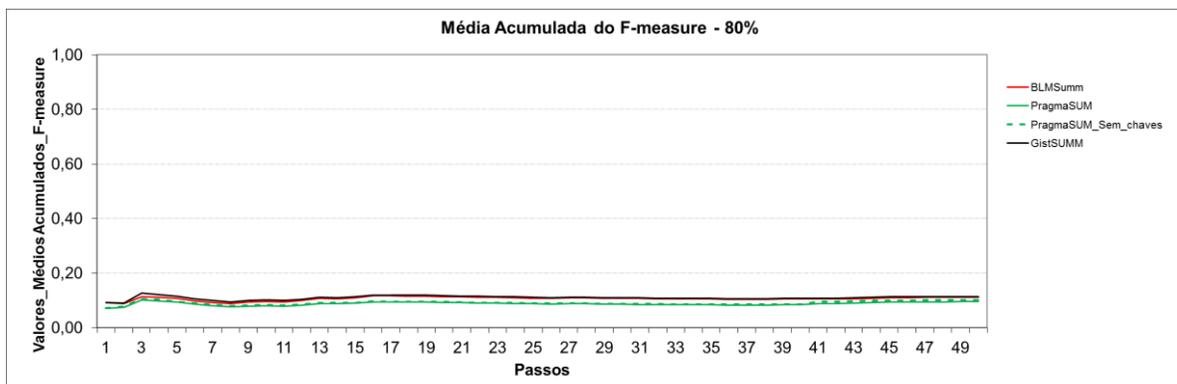


Figura 136. Média acumulada de F-Measure com taxa de compressão de 80% do domínio História.

A figura a seguir (média acumulada de Precision) mostram que, de uma maneira geral, as taxas de precisão alcançadas foram altas e que o sumário que obteve os melhores valores foi o PragmaSUM.

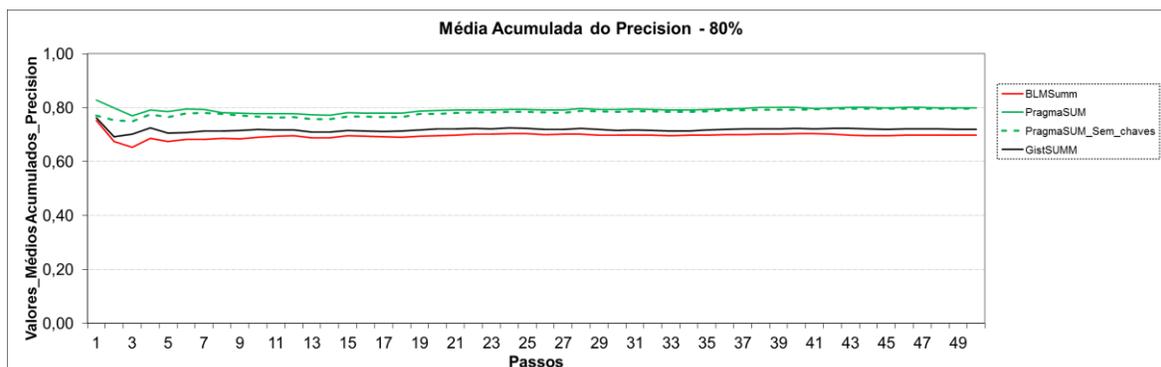


Figura 137. Média acumulada de Precision com taxa de compressão de 80% do domínio História.

A figura a seguir (média acumulada de Recall) evidenciam que há bastante equilíbrio nos resultados, com pouca variação e valores muito baixos.

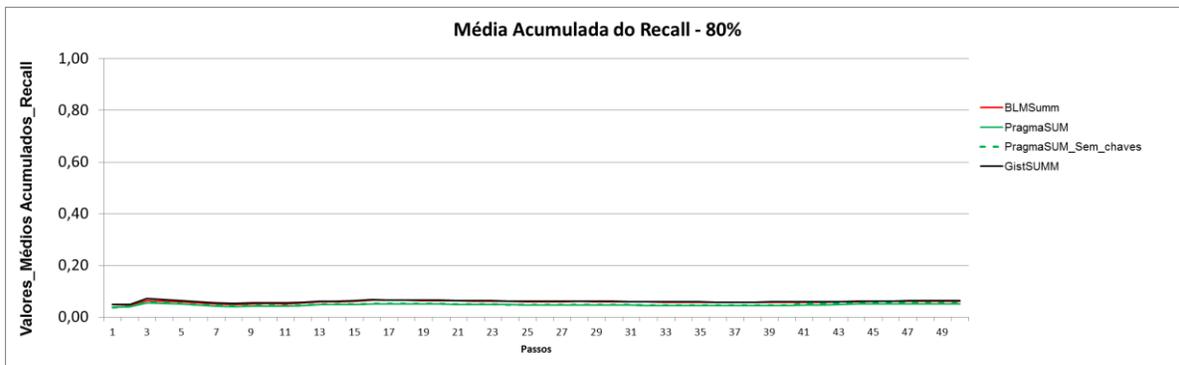


Figura 138. Média acumulada de Recall com taxa de compressão de 80% do domínio História.

8.4. Taxa de Compressão de 90%

Nas figuras desta seção têm-se os resultados das SA para os 50 textos do domínio História com taxa de compressão aplicada de 90%.

É possível perceber na figura a seguir (média acumulada de F-Measure) que, de forma geral, o BLMSumm apresenta os resultados mais satisfatórios junto ao GistSumm seguido proximamente pelo PragmaSUM, com pouca variação nos valores.

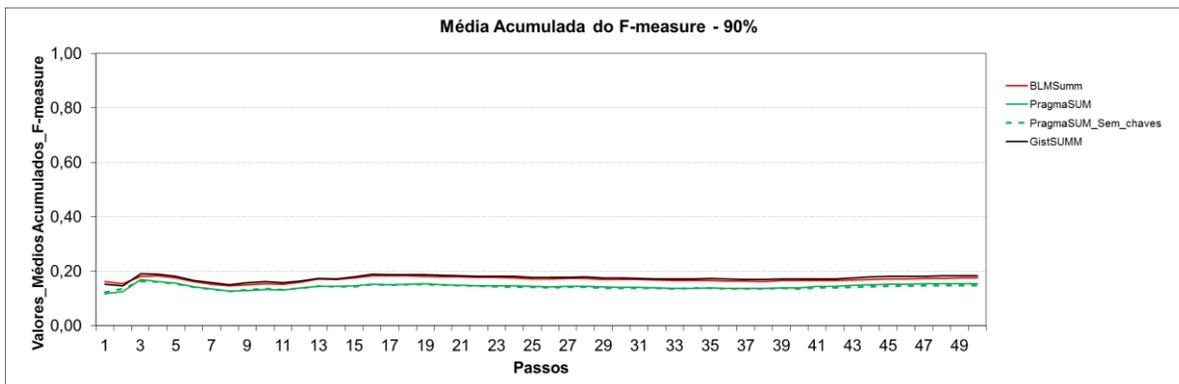


Figura 139. Média acumulada de F-Measure com taxa de compressão de 90% do domínio História.

É possível perceber na figura a seguir (média acumulada de Precision) que quanto à precisão o PragmaSUM obteve os melhores resultados dentre os sumarizadores analisados, com ligeira vantagem sobre sua versão sem chaves.

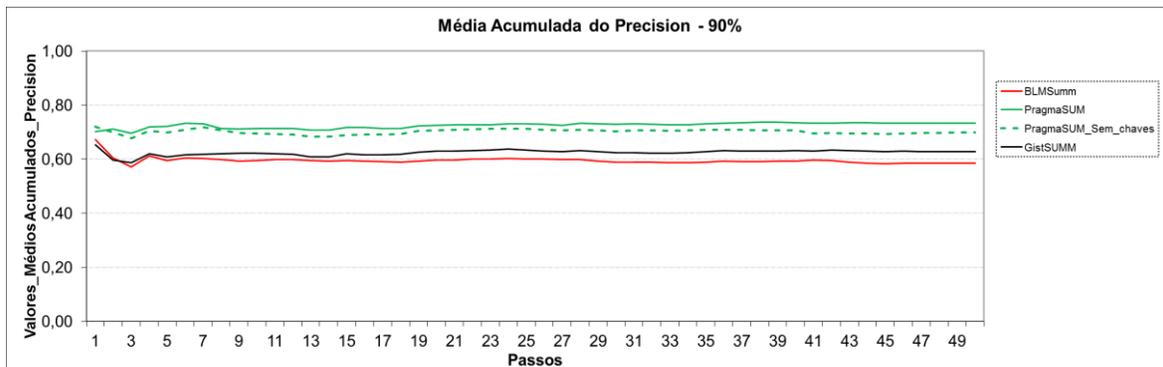


Figura 140. Média acumulada de Precision com taxa de compressão de 90% do domínio História.

É possível perceber na figura a seguir (média acumulada de Recall) que há bastante equilíbrio nos resultados, com pouca variação e valores muito baixos.

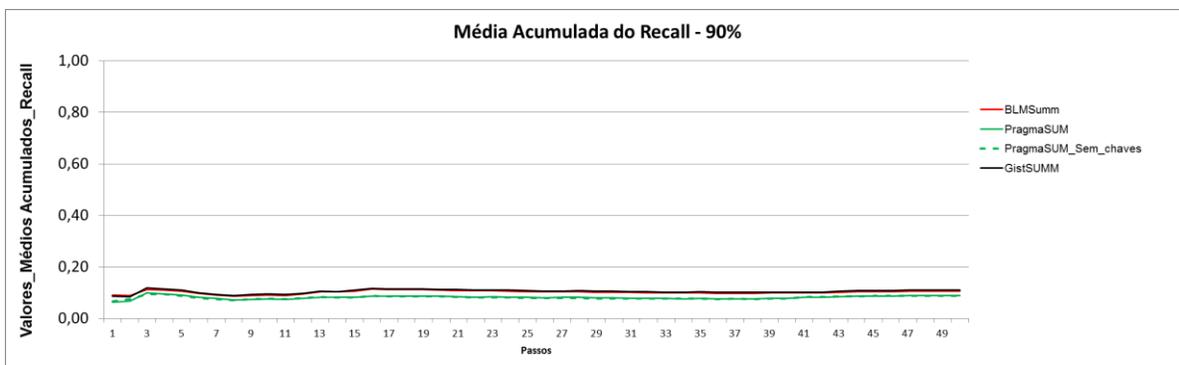


Figura 141. Média acumulada de Recall com taxa de compressão de 90% do domínio História.

8.5. Comparação dos Resultados

A tabela a seguir mostra uma comparação feita entre as taxas de compressão das médias gerais de média harmônica (F-Measure) alcançadas pelos sumarizadores automáticos para os textos do domínio História. Os valores variam pouco entre si, com ligeira vantagem ao GistSumm. A Figura 142 reafirma as observações.

Tabela 31. Comparação dos resultados de F-Measure do domínio História.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves

50%	0,05702	0,08129	0,05032	0,05237
70%	0,08403	0,08986	0,07175	0,07418
80%	0,11152	0,11424	0,09529	0,10032
90%	0,1744	0,18269	0,15404	0,14789

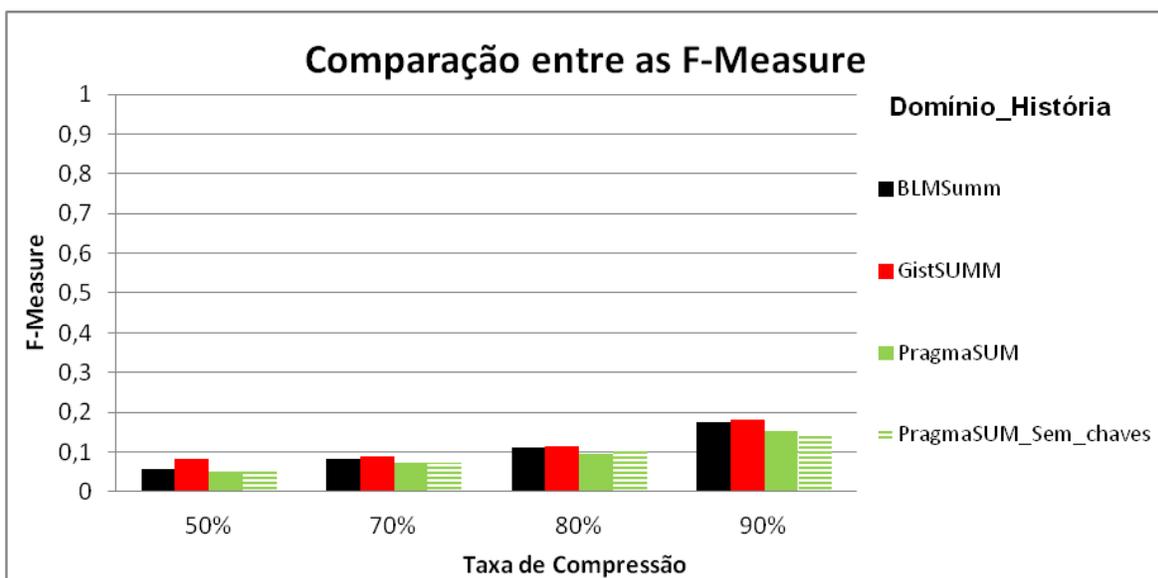


Figura 142. Comparação dos resultados de F-Measure do domínio História.

A tabela e figura a seguir trazem uma comparação feita entre as taxas de compressão das médias gerais de Recall (cobertura) que os sumarizadores automáticos alcançaram para os textos do domínio História. Constata-se que os valores ficaram muito próximos entre si.

Tabela 32. Comparação dos resultados de Recall do domínio História.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,02973	0,04333	0,02614	0,02727
70%	0,04508	0,04831	0,03799	0,03943
80%	0,06167	0,06314	0,05166	0,0553
90%	0,10638	0,1105	0,08894	0,08877

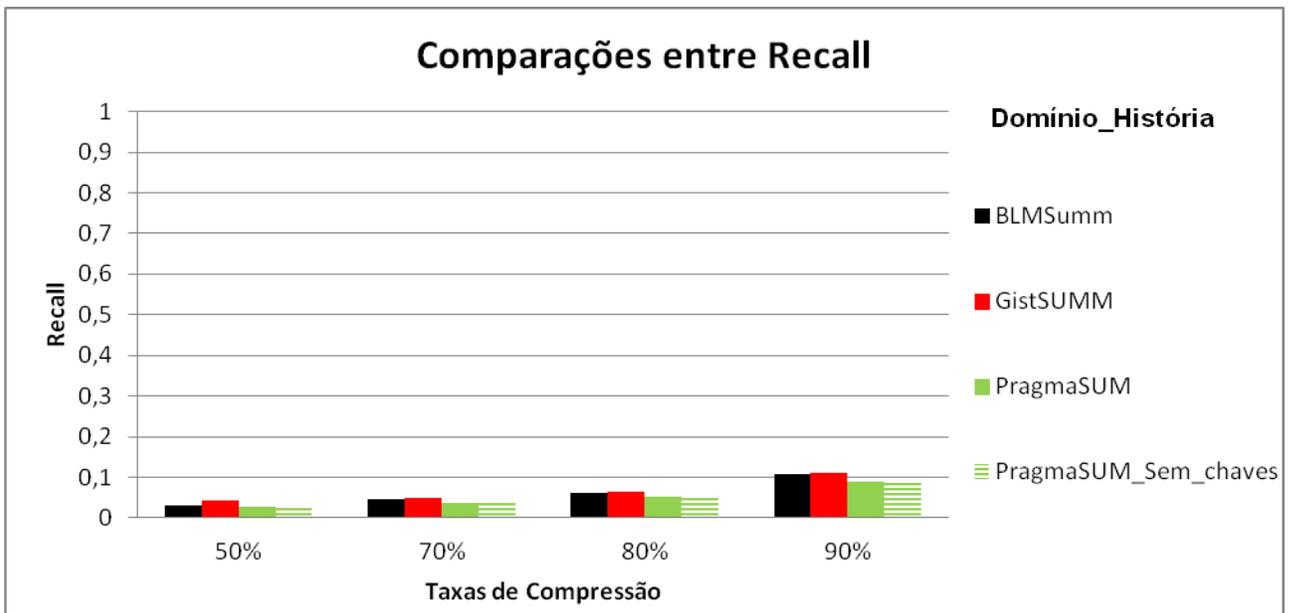


Figura 143. Comparação dos resultados de Recall do domínio História.

9. Domínio Medicina

Esta seção descreve os resultados das SA realizadas no domínio Medicina.

9.1. Taxa de Compressão de 50%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio Medicina. A taxa de compressão aplicada foi de 50%.

A figura a seguir (média acumulada de F-Measure) apresentam os resultados das SA para cada um dos sumarizadores automáticos. Observa-se que há bastante equilíbrio nos resultados, mas o GistSumm possui uma ligeira vantagem.

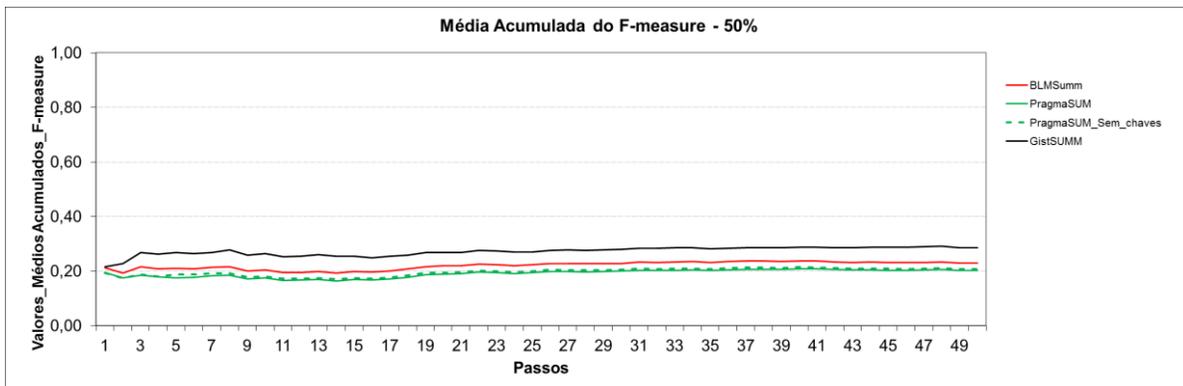


Figura 144. Média acumulada de F-Measure com taxa de compressão de 50% do domínio Medicina.

Na figura a seguir (média acumulada de Precision) têm-se os resultados para cada um dos sumarizadores automáticos. Quanto à precisão (Precision) o PragmaSUM apresentou os melhores resultados.

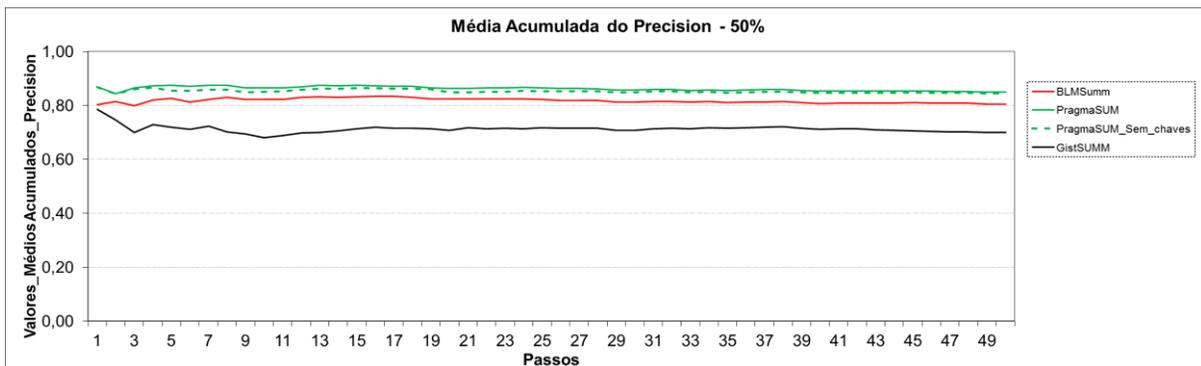


Figura 145. Média acumulada de Precision com taxa de compressão de 50% do domínio Medicina.

A figura a seguir (média acumulada de Recall) mostram os resultados da cobertura para cada um dos sumarizadores automáticos. É possível perceber que existe uma variação muito pequena dentre resultados aferidos e o GistSumm possui ligeira vantagem.

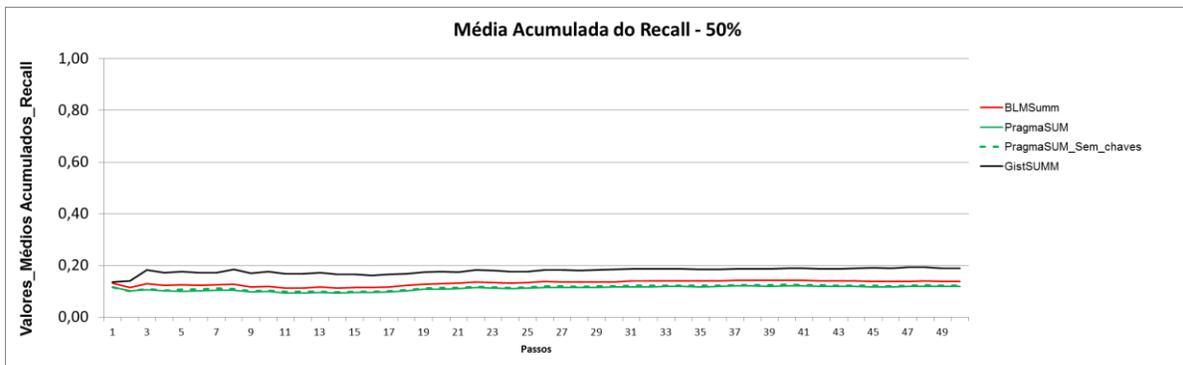


Figura 146. Média acumulada de Recall com taxa de compressão de 50% do domínio Medicina.

9.2. Taxa de Compressão de 70%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio Medicina para uma taxa de compressão de 70%.

É possível perceber na figura a seguir (média acumulada de F-Measure) que o BLMSumm apresenta os resultados mais satisfatórios junto com o GistSumm, com pouca variação entre os resultados.

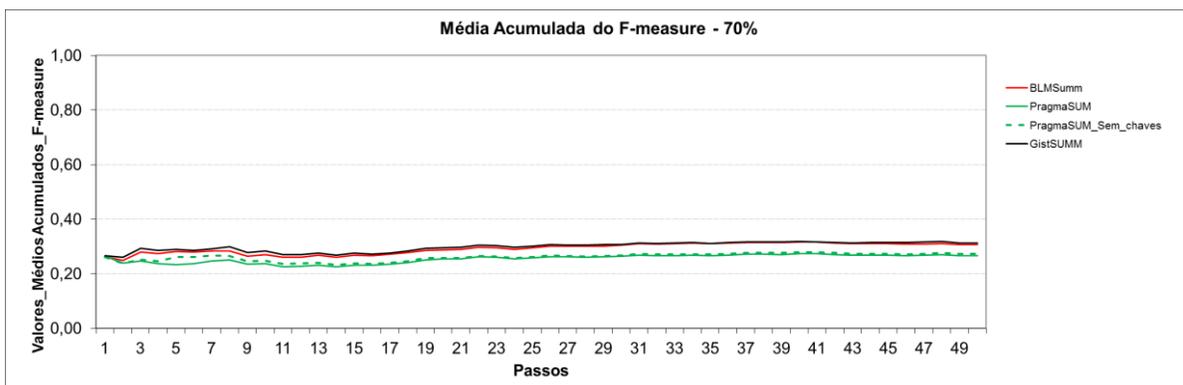


Figura 147. Média acumulada de F-Measure com taxa de compressão de 70% do domínio Medicina.

Percebe-se na figura a seguir (média acumulada de Precision) que o sumarizador com os valores de precisão mais altos foi o PragmaSUM e o com valores mais baixos foi o GistSumm.

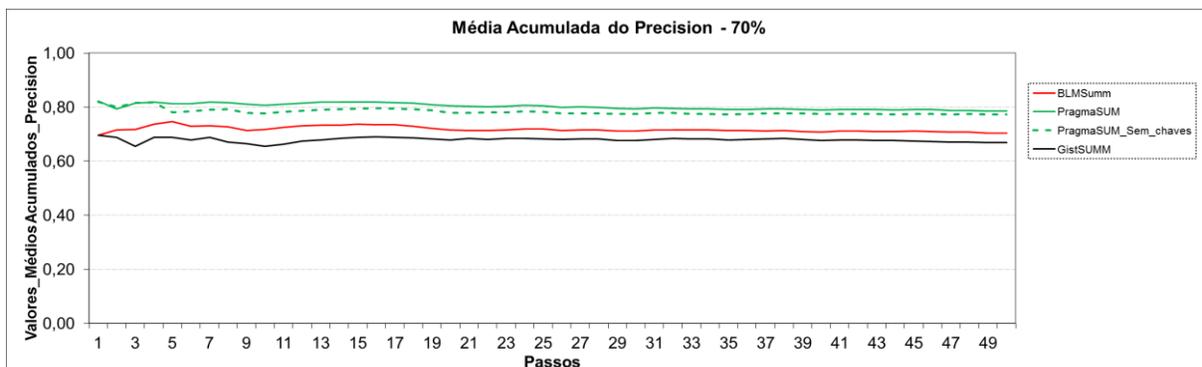


Figura 148. Média acumulada de Precision com taxa de compressão de 70% do domínio Medicina.

A figura a seguir (média acumulada de Recall) mostram que quanto à cobertura os sumarizadores obtiveram resultados baixos e bem próximos entre si.

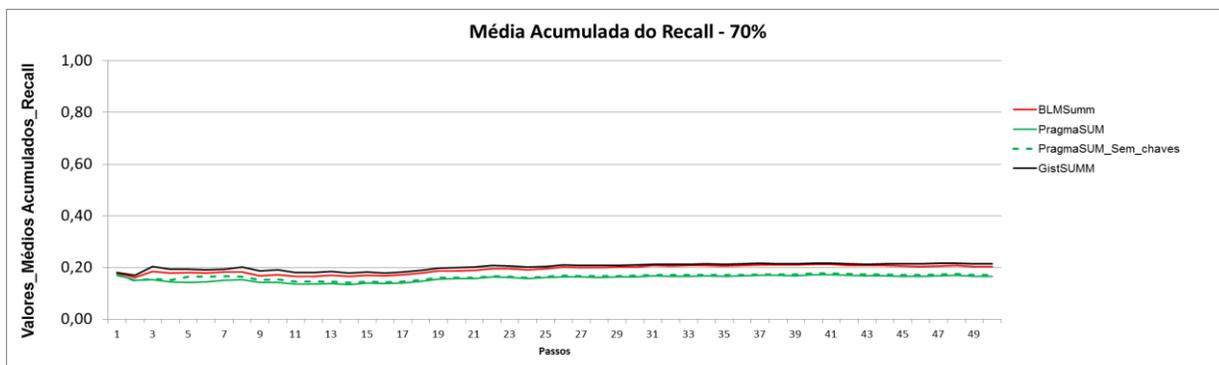


Figura 149. Média acumulada de Recall com taxa de compressão de 70% do domínio Medicina.

9.3. Taxa de Compressão de 80%

As figuras a seguir trazem os resultados das SA para os 50 textos do domínio Medicina com taxa de compressão de 80%.

Na figura a seguir (média acumulada de F-Measure) percebe-se que o resultados são muito próximos entre si e que o BLMSumm junto com o GistSumm possuem pequena vantagem.

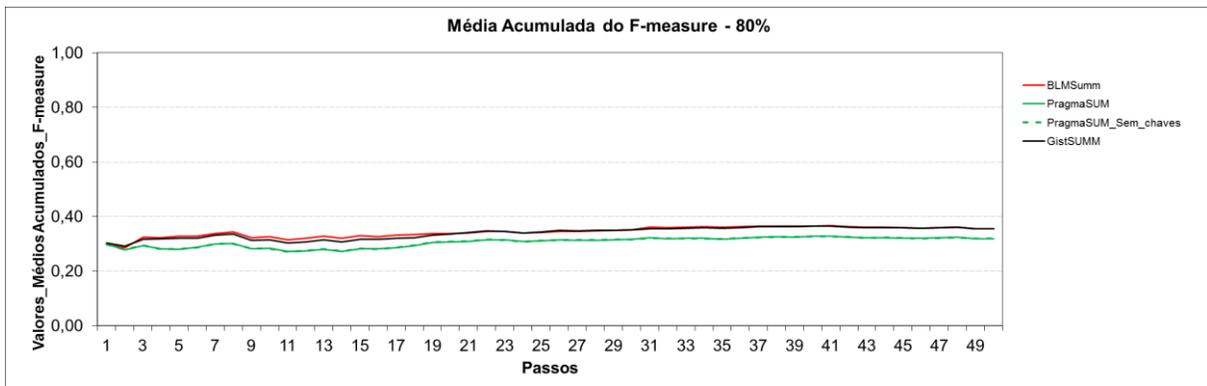


Figura 150. Média acumulada de F-Measure com taxa de compressão de 80% do domínio Medicina.

A figura a seguir (média acumulada de Precision) mostram que, de uma maneira geral, as taxas de precisão alcançadas foram altas e que o sumarizador que obteve os melhores valores foi o PragmaSUM.

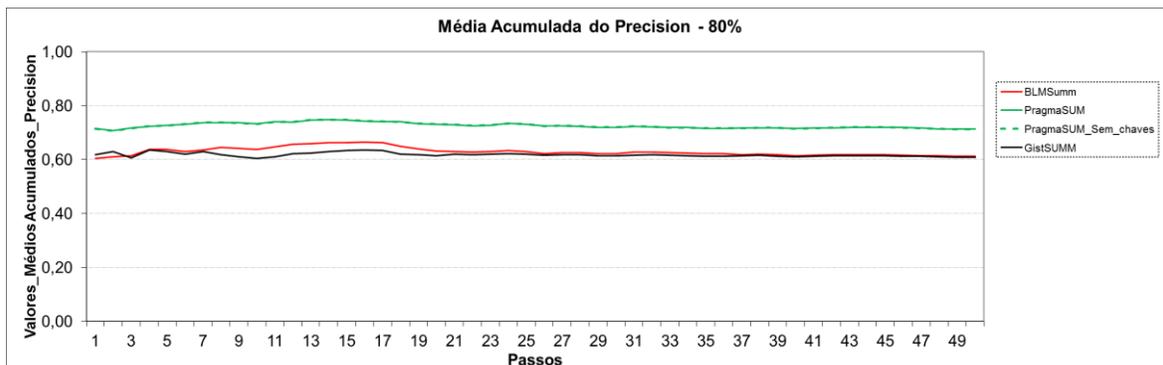


Figura 151. Média acumulada de Precision com taxa de compressão de 80% do domínio Medicina.

A figura a seguir (média acumulada de Recall) evidenciam que quanto à cobertura o BLMSumm junto ao GistSumm obtiveram ligeira vantagem, mas os valores variam pouco entre si.

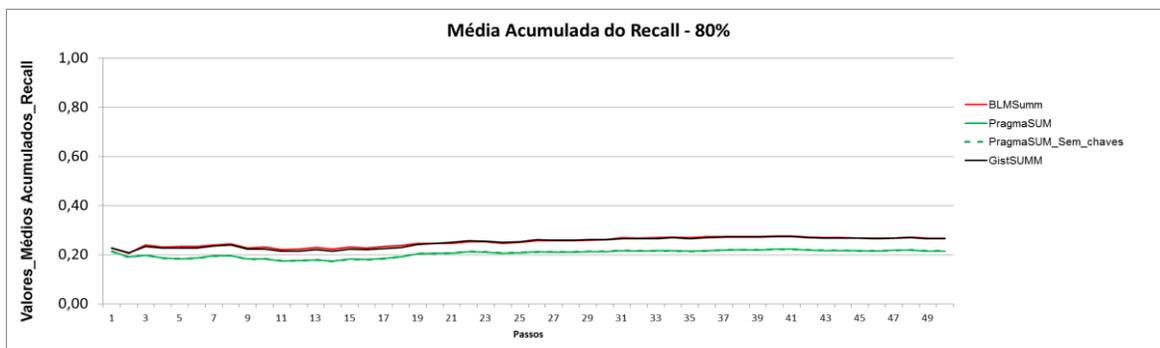


Figura 152. Média acumulada de Recall com taxa de compressão de 80% do domínio Medicina.

9.4. Taxa de Compressão de 90%

Nas figuras a seguir têm-se os resultados das SA para os 50 textos do domínio Medicina com taxa de compressão aplicada de 90%.

É possível perceber na figura a seguir (média acumulada de F-Measure) que, de forma geral, os resultados não apresentam muita variação e ficam muito próximos entre si.

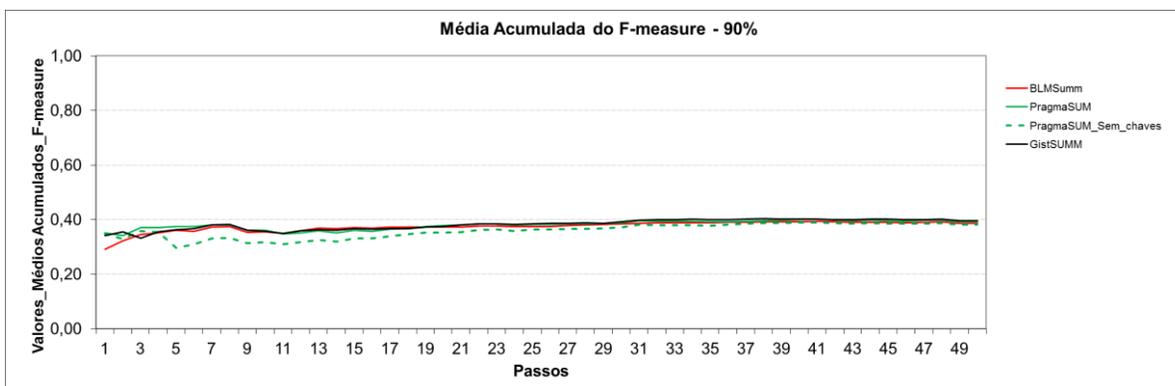


Figura 153. Média acumulada de F-Measure com taxa de compressão de 90% do domínio Medicina.

É possível perceber na figura a seguir (média acumulada de Precision) que quanto à precisão o PragmaSUM obteve os melhores resultados dentre os sumarizadores analisados, seguido por sua versão sem chaves. BLMSumm e GistSumm obtiveram valores muito próximos entre si.

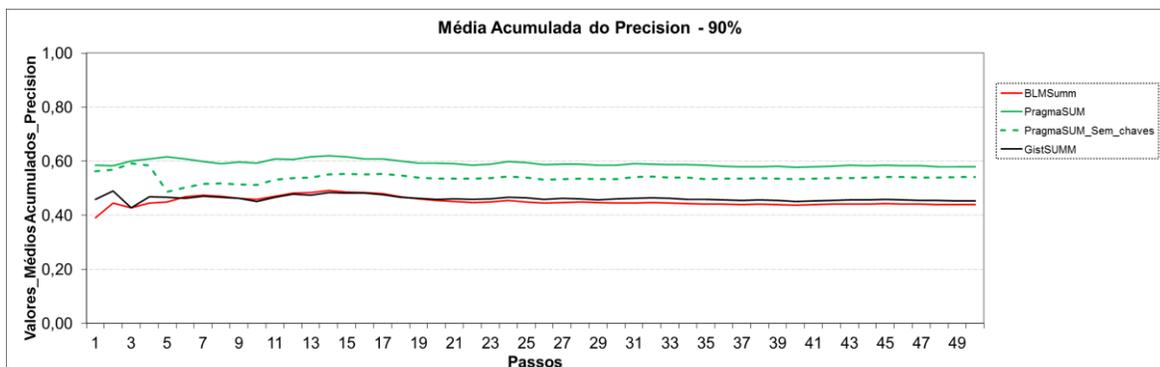


Figura 154. Média acumulada de Precision com taxa de compressão de 90% do domínio Medicina.

É possível perceber na figura a seguir (média acumulada de Recall) que quanto à cobertura o BLMSumm obteve os melhores resultados junto com o GistSumm com ligeira vantagem sobre o PragmaSUM.

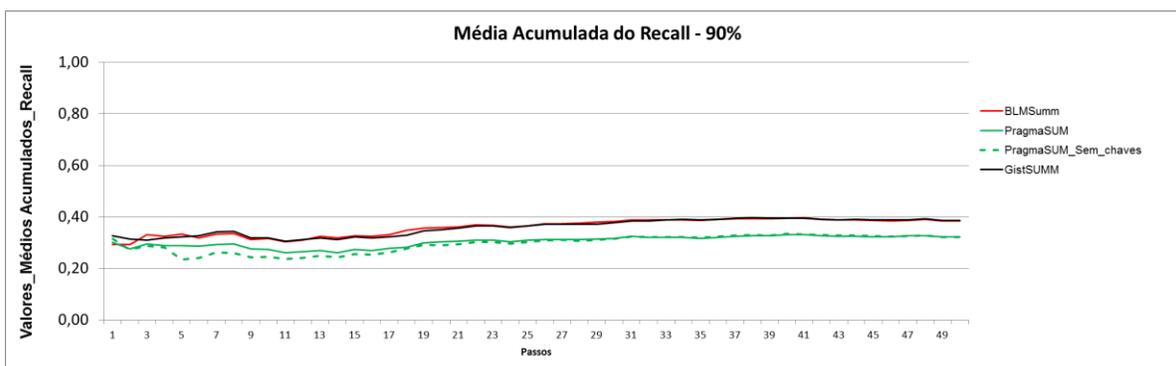


Figura 155. Média acumulada de Recall com taxa de compressão de 90% do domínio Medicina.

9.5. Comparação dos Resultados

A tabela a seguir mostra uma comparação feita entre as taxas de compressão das médias gerais de média harmônica (F-Measure) alcançadas pelos sumarizadores automáticos para os textos do domínio Medicina. Os resultados ficaram muito próximos entre si, exceto com 50% de compressão, onde o GistSumm obteve melhores resultados. A Figura 156 reafirma as observações.

Tabela 33. Comparação dos resultados de F-Measure do domínio Medicina.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,22911	0,2856	0,20233	0,20723
70%	0,30623	0,31198	0,26567	0,2715
80%	0,35577	0,35424	0,31798	0,31798
90%	0,38597	0,3958	0,39166	0,38146

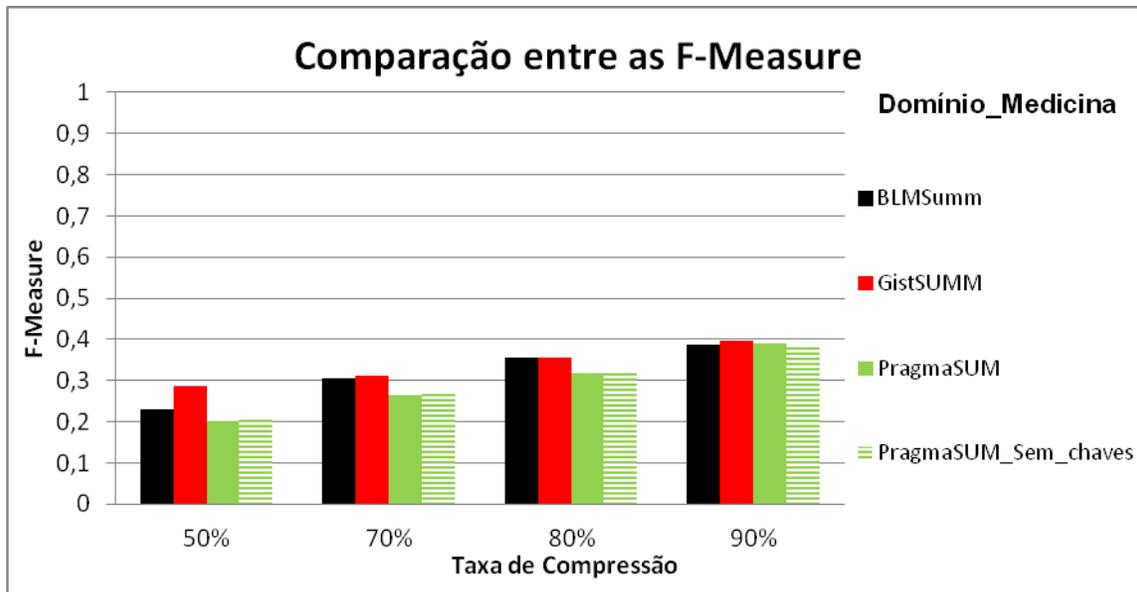


Figura 156. Comparação dos resultados de F-Measure do domínio Medicina.

A tabela e figura a seguir trazem uma comparação feita entre as taxas de compressão das médias gerais de Recall (cobertura) que os sumarizadores automáticos alcançaram para os textos do domínio Medicina. Constata-se que o GistSumm apresentou os melhores resultados para a taxa de 50% de compressão, mas depois é igualado pelo BLMSumm com valores muito próximos.

Tabela 34. Comparação dos resultados de Recall do domínio Medicina.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,13721	0,18889	0,11748	0,12096
70%	0,20389	0,21328	0,16548	0,1708
80%	0,26607	0,26432	0,21479	0,21479
90%	0,38376	0,38559	0,32143	0,32153

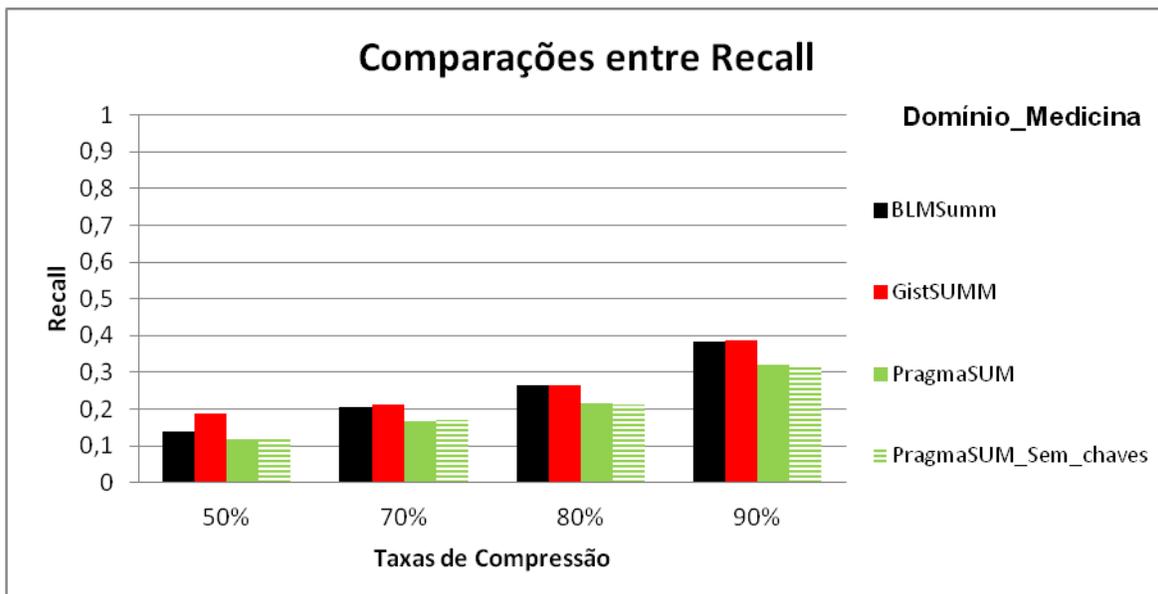


Figura 157. Comparação dos resultados de Recall do domínio Medicina.

10. Domínio Psicologia

Esta seção descreve os resultados das SA realizadas no domínio Psicologia.

10.1. Taxa de Compressão de 50%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio Psicologia. A taxa de compressão aplicada foi de 50%.

A figura a seguir (média acumulada de F-Measure) apresentam os resultados das SA para cada um dos sumarizadores automáticos. Observa-se que há bastante equilíbrio nos resultados, mas o GistSumm possui uma ligeira vantagem.

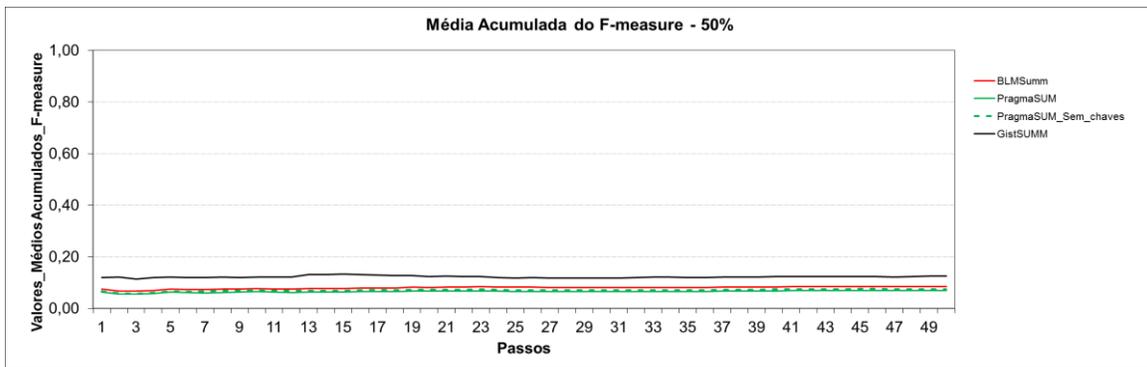


Figura 158. Média acumulada de F-Measure com taxa de compressão de 50% do domínio Psicologia.

Na figura a seguir (média acumulada de Precision) têm-se os resultados para cada um dos sumarizadores automáticos. Quanto à precisão (Precision) o PragmaSUM apresentou os melhores resultados.

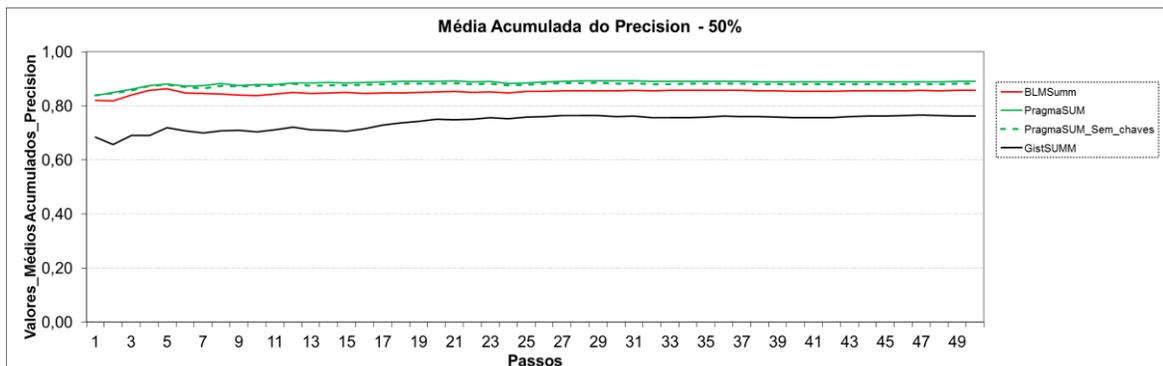


Figura 159. Média acumulada de Precision com taxa de compressão de 50% do domínio Psicologia.

A figura a seguir (média acumulada de Recall) mostram os resultados da cobertura para cada um dos sumarizadores automáticos. É possível perceber que existe uma variação muito pequena dentre resultados aferidos e o GistSumm possui ligeira vantagem.

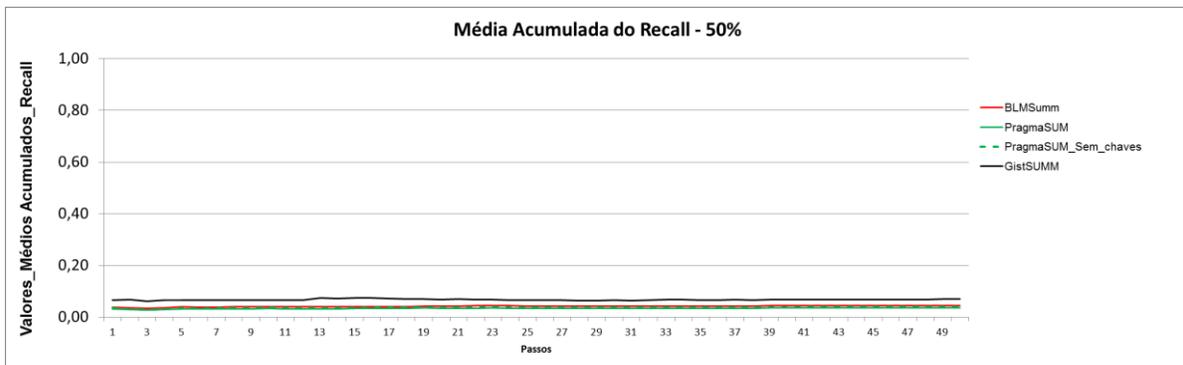


Figura 160. Média acumulada de Recall com taxa de compressão de 50% do domínio Psicologia.

10.2. Taxa de Compressão de 70%

As figuras desta seção trazem os resultados das SA para os 50 textos do domínio Psicologia para uma taxa de compressão de 70%.

É possível perceber na figura a seguir (média acumulada de F-Measure) que é possível perceber que existe uma variação muito pequena dentre resultados aferidos e o GistSumm possui ligeira vantagem.

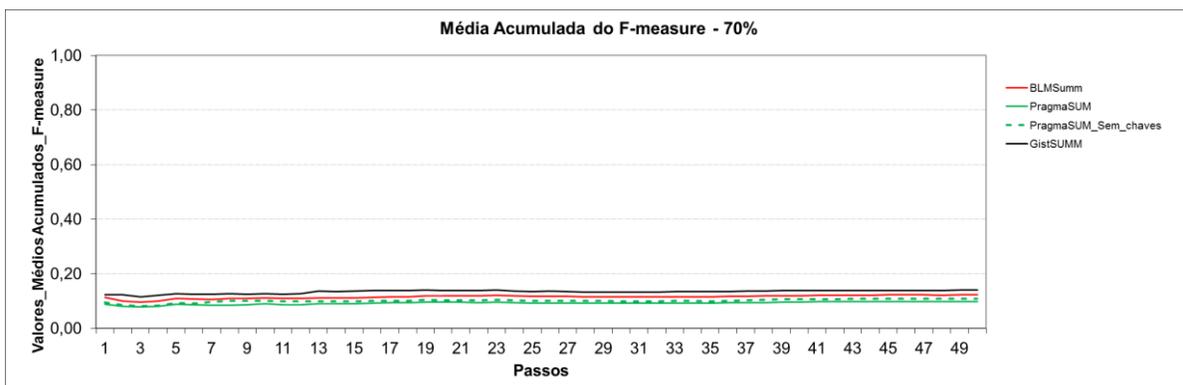


Figura 161. Média acumulada de F-Measure com taxa de compressão de 70% do domínio Psicologia.

Percebe-se na figura a seguir (média acumulada de Precision) que o sumarizador com os valores de precisão mais altos foi o PragmaSUM, com ligeira vantagem sobre sua versão sem chaves.

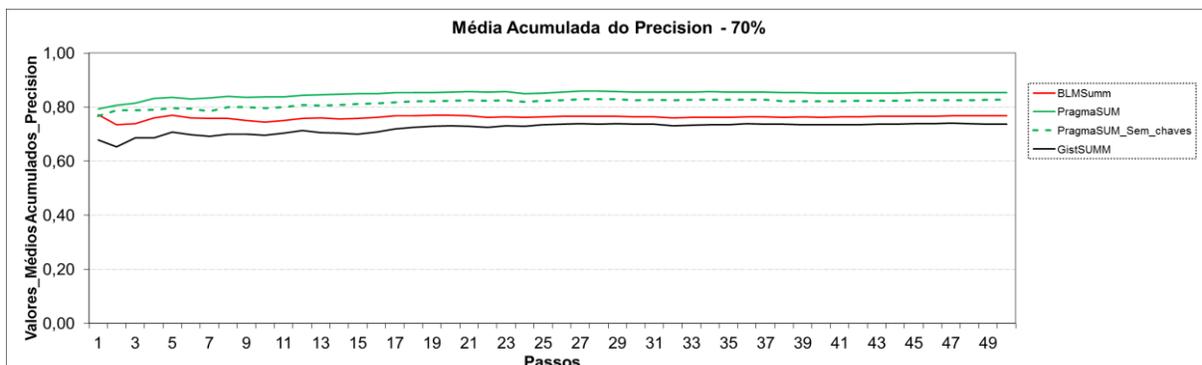


Figura 162. Média acumulada de Precision com taxa de compressão de 70% do domínio Psicologia.

A figura a seguir (média acumulada de Recall) mostram que quanto à cobertura os sumarizadores obtiveram resultados similares e sem muita variação entre si.

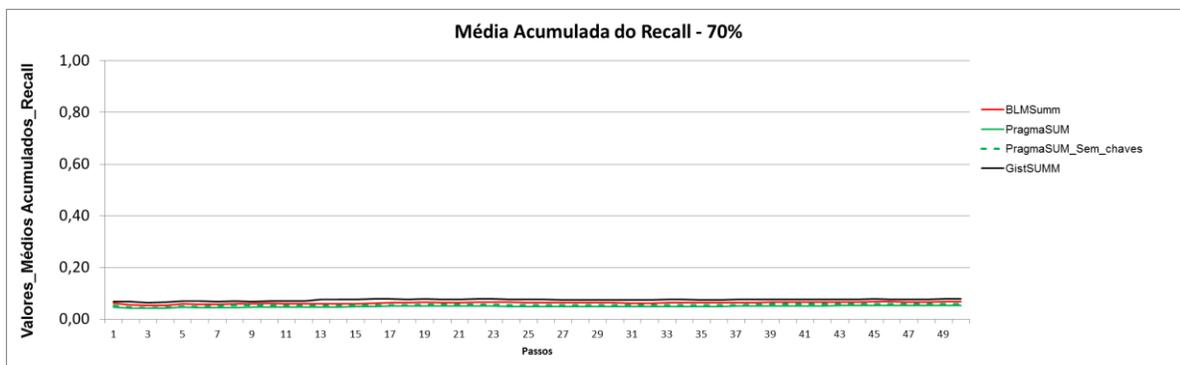


Figura 163. Média acumulada de Recall com taxa de compressão de 70% do domínio Psicologia.

10.3. Taxa de Compressão de 80%

As tabelas figuras desta seção trazem os resultados das SA para os 50 textos do domínio Psicologia com taxa de compressão de 80%.

Na figura a seguir (média acumulada de F-Measure) percebe-se que o BLMSumm apresenta os resultados mais satisfatórios junto com o GistSumm com uma ligeira vantagem sobre o PragmaSUM.

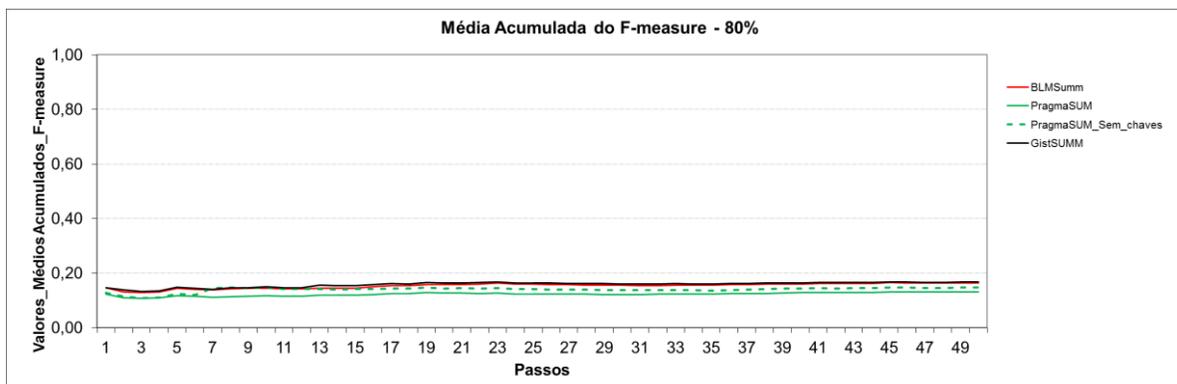


Figura 164. Média acumulada de F-Measure com taxa de compressão de 80% do domínio Psicologia.

A figura a seguir (média acumulada de Precision) mostram que, de uma maneira geral, as taxas de precisão alcançadas foram altas e que o sumarizador que obteve os melhores valores foi o PragmaSUM seguido por sua versão sem chaves.

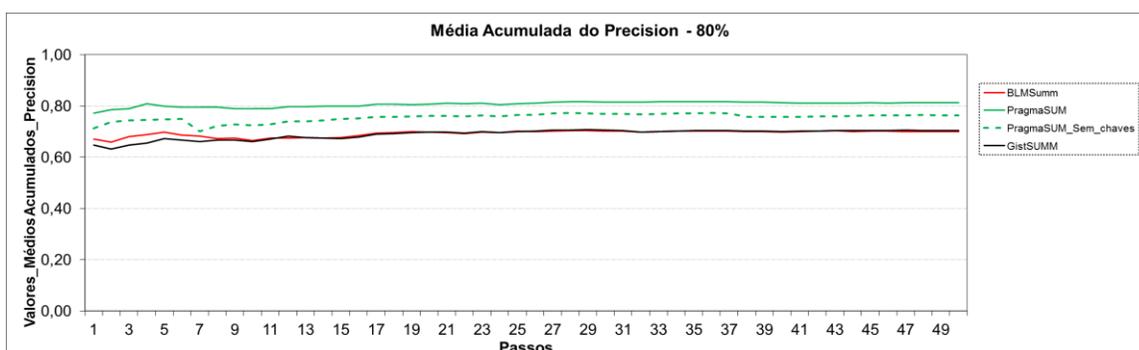


Figura 165. Média acumulada de Precision com taxa de compressão de 80% do domínio Psicologia.

A figura a seguir (média acumulada de Recall) evidenciam que quanto à cobertura os resultados foram homogêneos, apresentado pouca variação entre si.

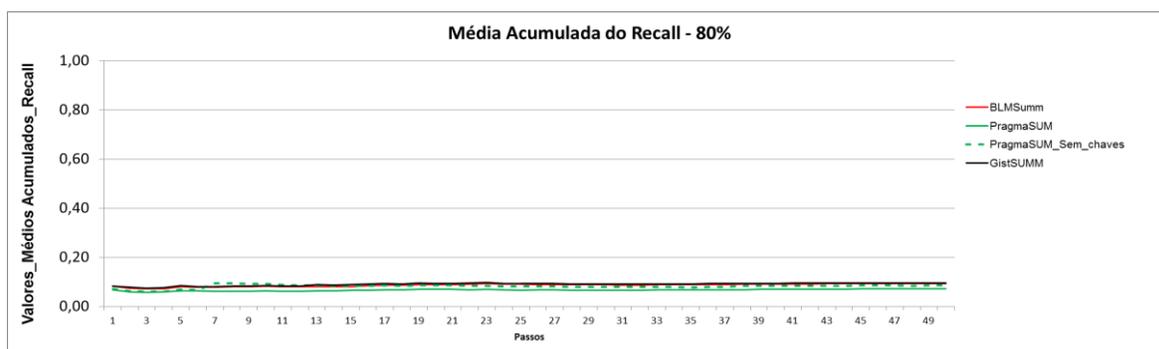


Figura 166. Média acumulada de Recall com taxa de compressão de 80% do domínio Psicologia.

10.4. Taxa de Compressão de 90%

Nas figuras desta seção têm-se os resultados das SA para os 50 textos do domínio Psicologia com taxa de compressão aplicada de 90%.

É possível perceber na figura a seguir (média acumulada de F-Measure) que, de forma geral, o BLMSumm apresenta os resultados mais satisfatórios junto com o GistSumm seguido proximamente pela versão sem chaves do PragmaSUM.

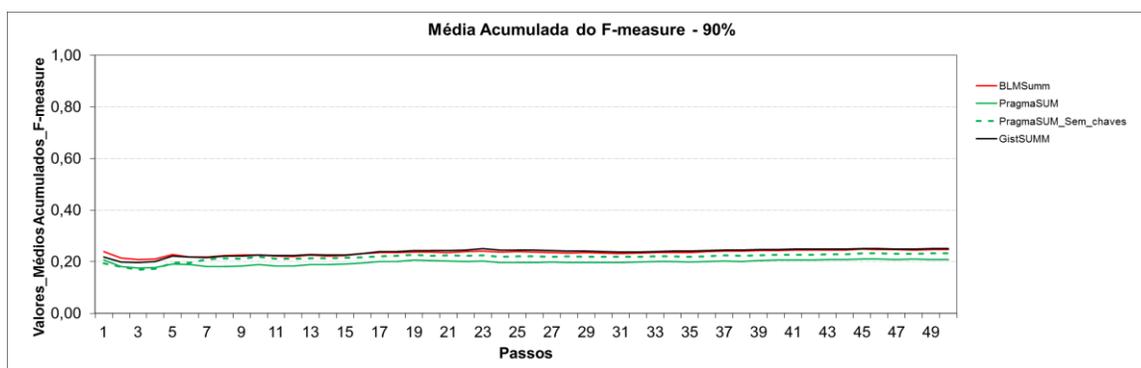


Figura 167. Média acumulada de F-Measure com taxa de compressão de 90% do domínio Psicologia.

É possível perceber na figura a seguir (média acumulada de Precision) que quanto à precisão o PragmaSUM obteve os melhores resultados dentre os sumarizadores analisados, com boa vantagem sobre sua versão sem chaves. BLMSumm e GistSumm obtiveram valores próximos entre si.

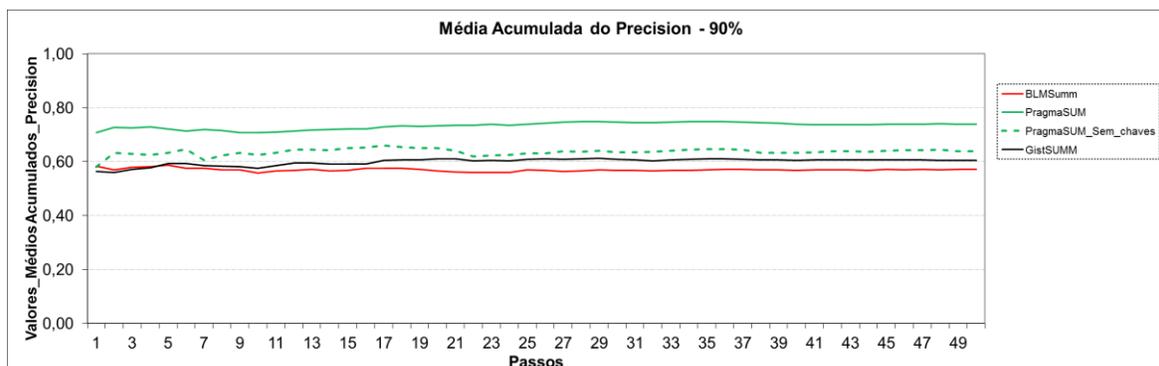


Figura 168. Média acumulada de Precision com taxa de compressão de 90% do domínio Psicologia.

É possível perceber na figura a seguir (média acumulada de Recall) que quanto à cobertura os resultados foram muito próximos entre si.

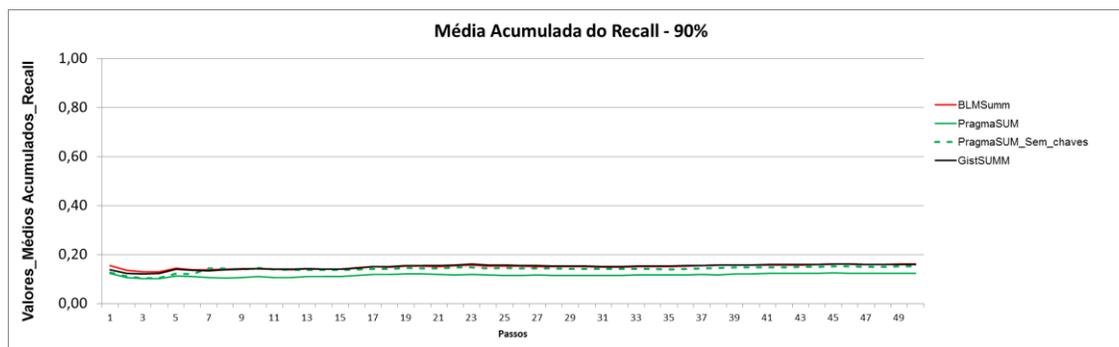


Figura 169. Média acumulada de Recall com taxa de compressão de 90% do domínio Psicologia.

10.5. Comparação dos Resultados

A tabela a seguir mostra uma comparação feita entre as taxas de compressão das médias gerais de média harmônica (F-Measure) alcançadas pelos sumarizadores automáticos para os textos do domínio Psicologia. O GistSumm obteve os melhores resultados com 50 e 70%, quando a taxa de compressão foi de 80 e 90% o BLMSumm se igualou a ele. Quando a taxa de compressão é de 90% os valores variam pouco entre si, e a versão sem chaves do PragmaSUM tem ligeira vantagem sobre sua versão que utiliza as chaves. A Figura 170 reafirma as observações.

Tabela 35. Comparação dos resultados de F-Measure do domínio Psicologia.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,08481	0,12519	0,06923	0,07443
70%	0,12288	0,14063	0,09839	0,10888
80%	0,16444	0,16785	0,13073	0,14668
90%	0,24757	0,25031	0,20916	0,23216

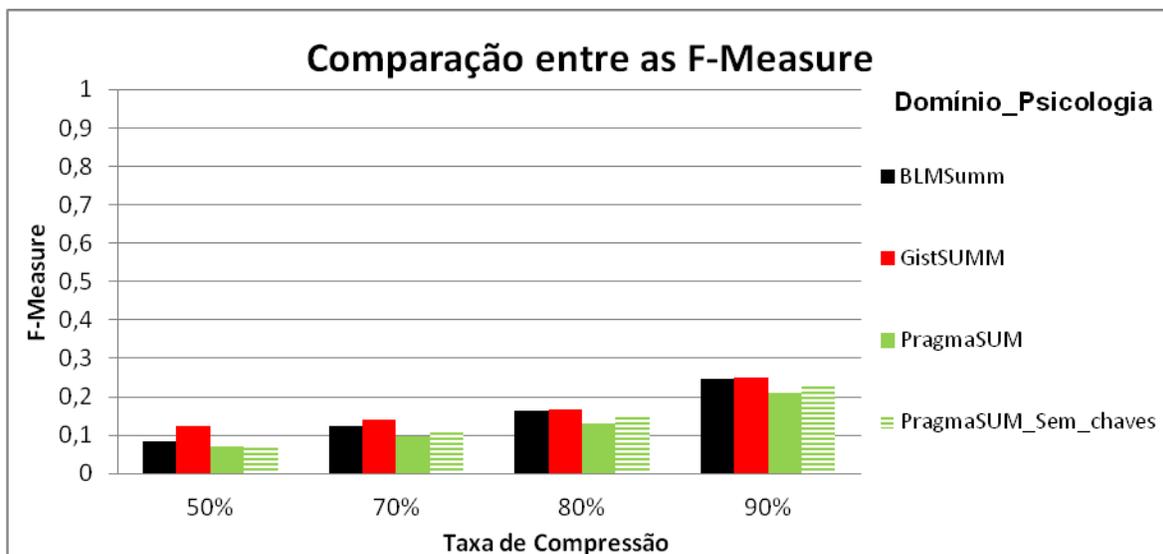


Figura 170. Comparação dos resultados de F-Measure do domínio Psicologia.

A tabela e figura a seguir trazem uma comparação feita entre as taxas de compressão das médias gerais de Recall (cobertura) que os sumarizadores automáticos alcançaram para os textos do domínio Psicologia. Constata-se que o BLMSumm apresentou os melhores resultados junto com o GistSumm com ligeira vantagem sobre o PragmaSUM. Os valores variam pouco entre si.

Tabela 36. Comparação dos resultados de Recall do domínio Psicologia.

	BLMSumm	GistSUMM	PragmaSUM	PragmaSUM_Sem_chaves
50%	0,04478	0,0696	0,0361	0,03896
70%	0,06712	0,0787	0,05241	0,05878
80%	0,09383	0,09619	0,07157	0,08508
90%	0,1605	0,16018	0,12355	0,15202

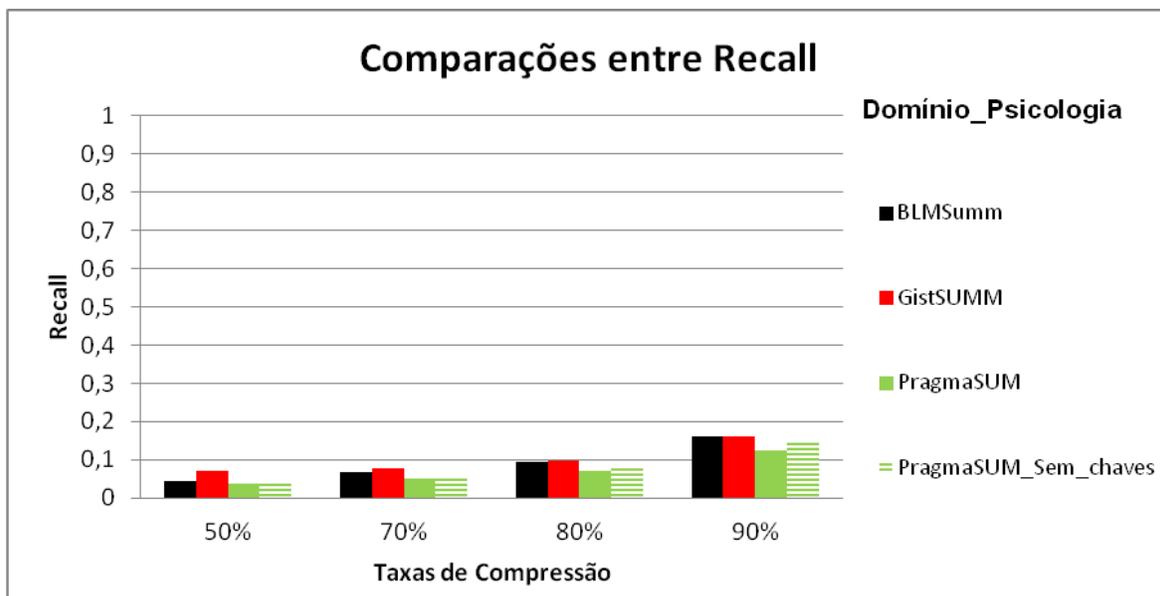


Figura 171. Comparação dos resultados de Recall do domínio Psicologia.

11. Comparações entre os domínios

Esta seção compara os resultados das sumarizações automáticas entre todos os domínios para as quatro taxas de compressão aplicadas.

11.1. Taxa de compressão de 50%

Para taxa de compressão de 50% o domínio Fisioterapia obteve os melhores resultados de média harmônica (F-Measure) como se pode observar na tabela e figura a seguir.

Tabela 37. Comparação da F-Measure para taxa de compressão de 50%.

	BLMSumm	GistSumm	PragmaSUM	PragmaSUM_Sem_chaves
Administração	0,10357	0,15867	0,0841	0,08541
Agronomia	0,24347	0,31408	0,19458	0,19702
Educação				
Física	0,10605	0,15514	0,08157	0,08347
Engenharia	0,10895	0,19551	0,08617	0,08121
Filosofia	0,06694	0,08954	0,0527	0,05763

Fisioterapia	0,25595	0,34934	0,21926	0,22816
Geografia	0,13934	0,19486	0,11686	0,12211
História	0,05702	0,08129	0,05032	0,05237
Medicina	0,22911	0,2856	0,20233	0,20723
Psicologia	0,08481	0,12519	0,06923	0,07443

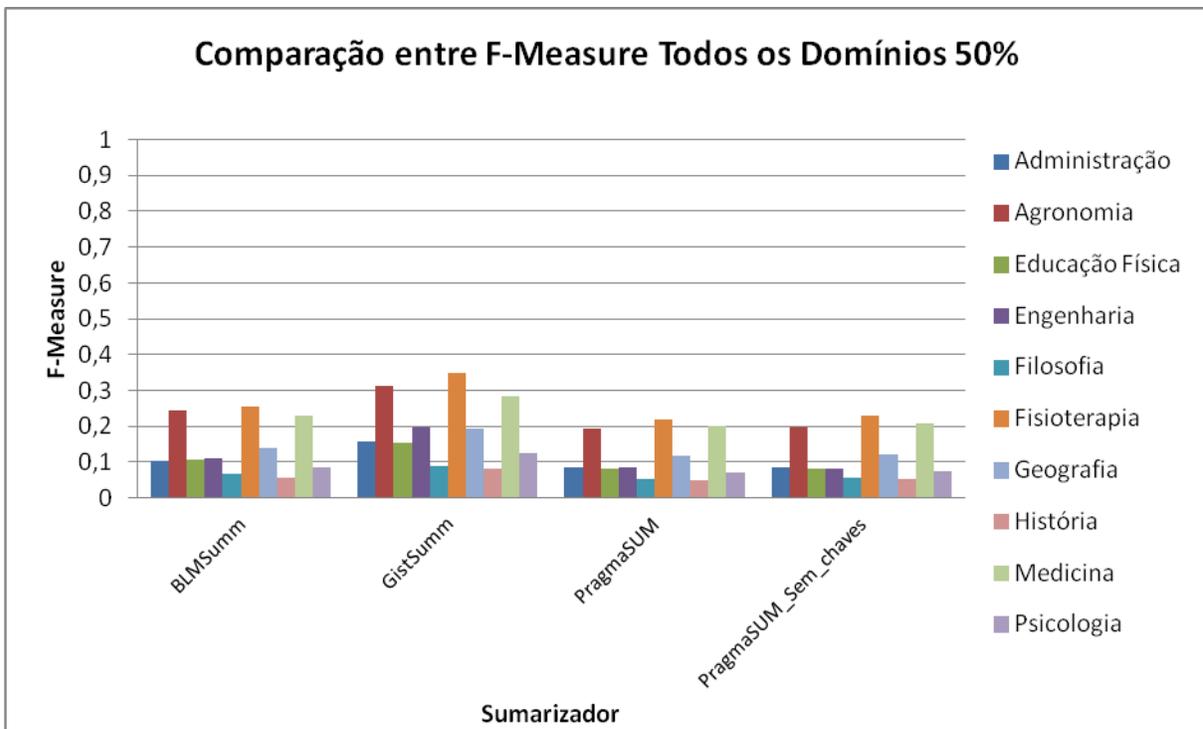


Figura 172. Comparação da F-Measure para taxa de compressão de 50%.

11.2. Taxa de compressão de 70%

Na figura a seguir se observa que quando a taxa de compressão aplicada foi 70% o domínio Fisioterapia obteve os melhores resultados de média harmônica (F-Measure) para todos sumarizadores.

Tabela 38. Comparação da F-Measure para taxa de compressão de 70%.

	BLMSumm	GistSumm	PragmaSUM	PragmaSUM_Sem_chaves
Administração	0,15311	0,16752	0,12017	0,1235
Agronomia	0,3211	0,32286	0,25297	0,26528
Educação Física	0,15071	0,16266	0,11199	0,11634
Engenharia	0,16302	0,19854	0,12369	0,11346
Filosofia	0,09914	0,10045	0,07409	0,07737
Fisioterapia	0,33299	0,35215	0,28551	0,30091
Geografia	0,19488	0,20866	0,16263	0,16773
História	0,08403	0,08986	0,07175	0,07418
Medicina	0,30623	0,31198	0,26567	0,2715
Psicologia	0,12288	0,14063	0,09839	0,10888

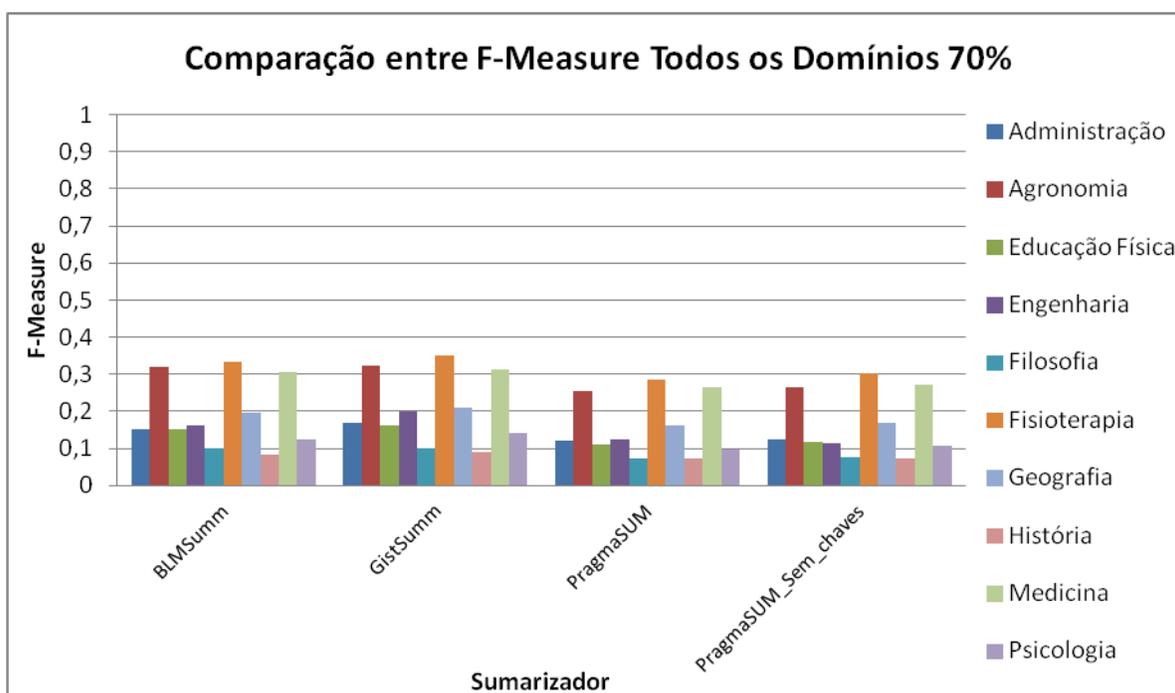


Figura 173. Comparação da F-Measure para taxa de compressão de 70%.

11.3. Taxa de compressão de 80%

A tabela e figura a seguir mostram que o domínio Fisioterapia com taxa de compressão de 80% auferiu resultados mais satisfatórios para os sumarizadores avaliados.

Tabela 39. Comparação da F-Measure entre os domínios médico e jornalístico para taxa de compressão de 80%.

	BLMSumm	GistSumm	PragmaSUM	PragmaSUM_Sem_chaves
Administração	0,20261	0,20153	0,15959	0,16485
Agronomia	0,37838	0,35833	0,30593	0,31793
Educação Física	0,19394	0,18859	0,14482	0,14811
Engenharia	0,21243	0,2162	0,16364	0,14849
Filosofia	0,13124	0,12781	0,0988	0,09933
Fisioterapia	0,39141	0,38723	0,34163	0,3436
Geografia	0,24345	0,24161	0,20539	0,21104
História	0,11152	0,11424	0,09529	0,10032
Medicina	0,35577	0,35424	0,31798	0,31798
Psicologia	0,16444	0,16785	0,13073	0,14668

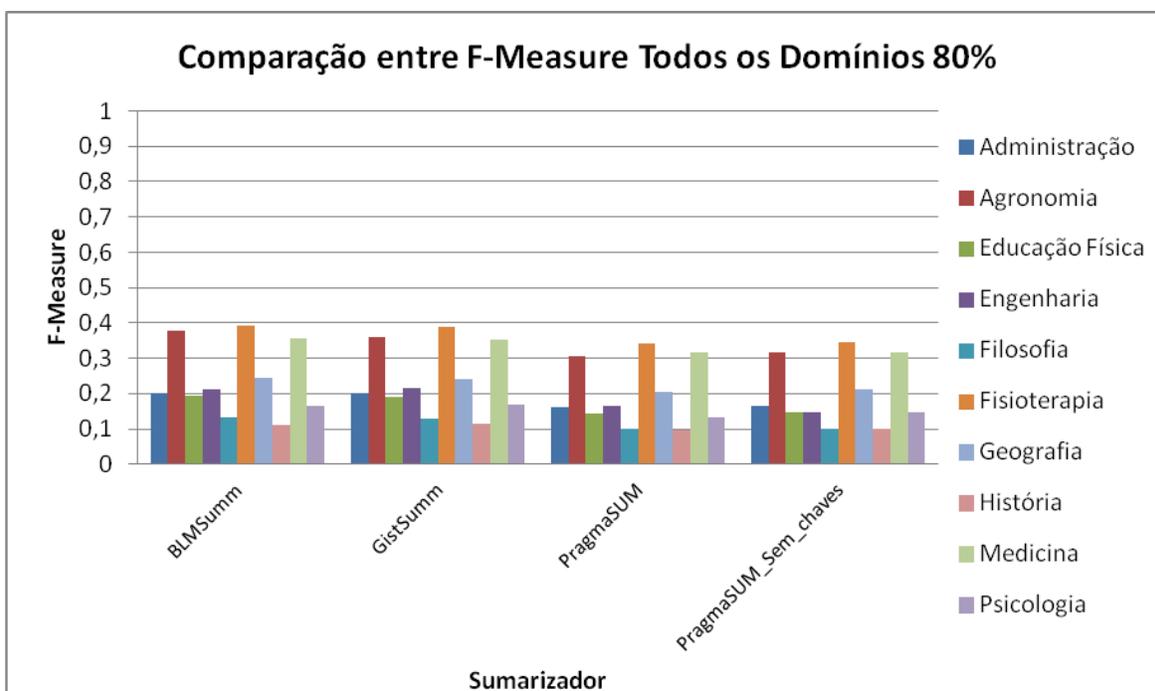


Figura 174. Comparação da F-Measure para taxa de compressão de 80%.

11.4. Taxa de compressão de 90%

Na tabela e figura a seguir percebe-se que para taxa de compressão de 90% o domínio Fisioterapia auferiu resultados mais satisfatórios com os sumarizadores utilizados.

Tabela 40. Comparação da F-Measure para taxa de compressão de 90%.

	BLMSumm	GistSumm	PragmaSUM	PragmaSUM_Sem_chaves
Administração	0,29702	0,29819	0,24682	0,24732
Agronomia	0,42329	0,40854	0,39453	0,38275
Educação Física	0,26947	0,2577	0,21994	0,21181
Engenharia	0,30255	0,29965	0,24917	0,23076
Filosofia	0,19841	0,1931	0,1539	0,15318
Fisioterapia	0,41891	0,42016	0,42006	0,39384
Geografia	0,31793	0,3162	0,2888	0,29406
História	0,1744	0,18269	0,15404	0,14789
Medicina	0,38597	0,3958	0,39166	0,38146
Psicologia	0,24757	0,25031	0,20916	0,23216

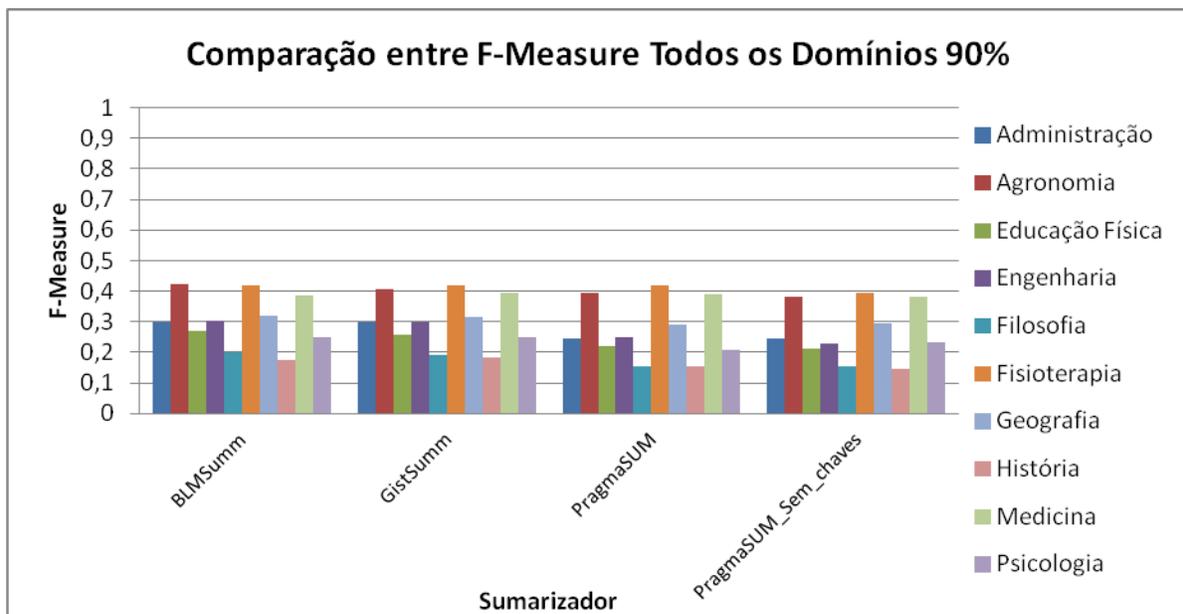


Figura 175. Comparação da F-Measure para taxa de compressão de 90%.

APÊNDICE B – TABELAS COM OS RESULTADOS DOS TESTES ESTATÍSTICOS DO F-MEASURE, PRECISION E RECALL

1. Domínio Administração

Tabela 41. Teste estatístico da métrica F-Measure – Domínio Administração.

Comparando amostras múltiplas relacionadas – F-Measure – Domínio Administração												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	150	p-nível	0	147,696	p-nível	0	150	p-nível	0	123,48	p-nível	0
Coef. de concordância de Kendall	1	Ordem médio	1	0,98464	Ordem médio	0,984327	1	Ordem médio	1	0,8232	Ordem médio	0,819592
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	3	150	0,109838	3,04	152	0,161933	4	200	0,213752	3,48	174	0,311713
PragmaSUM	1	50	0,088012	1	50	0,126209	1	50	0,167723	1,26	63	0,259178
PragmaSUM_S em_chaves	2	100	0,090491	2	100	0,130477	2	100	0,176214	1,74	87	0,263754
GistSUMM	4	200	0,16801	3,96	198	0,176061	3	150	0,210334	3,52	176	0,311953

Tabela 42. Teste estatístico da métrica Precision – Domínio Administração.

Comparando amostras múltiplas relacionadas – Precision – Domínio Administração												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3	50	Graus de liberdade	3	50	Graus de liberdade	3	50	Graus de liberdade	3
qui-quadrado	150	p-nível	0	150	p-nível	0	146,616	p-nível	0	148,824	p-nível	0
Coef. de concordância de Kendall	1	Ordem médio	1	1	Ordem médio	1	0,97744	Ordem médio	0,976979592	0,99216	Ordem médio	0,992
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	2	100	0,895922	2	100	0,826062813	1,94	97	0,75938549	1,02	51	0,636425656
PragmaSUM	4	200	0,920436	4	200	0,886554631	4	200	0,845178742	4	200	0,767893322
PragmaSUM_S em_chaves	3	150	0,913271	3	150	0,870993834	3	150	0,828541815	3	150	0,744961495
GistSUMM	1	50	0,797068	1	50	0,786091709	1,06	53	0,74968856	1,98	99	0,647740347

Tabela 43. Teste estatístico da métrica Recall – Domínio Administração.

Comparando amostras múltiplas relacionadas – Recall – Domínio Administração												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	150	p-nível	0	147,696	p-nível	0	150	p-nível	0	141,936	p-nível	0
Coef. de concordância de Kendall	1	Ordem médio	1	0,98464	Ordem médio	0,984327	1	Ordem médio	1	0,94624	Ordem médio	0,945143
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	3	150	0,05906	3,04	152	0,091162	4	200	0,127207	3,84	192	0,214665
PragmaSUM	1	50	0,046522	1	50	0,068638	1	50	0,094423	1	50	0,159896
PragmaSUM_S em_chaves	2	100	0,047954	2	100	0,071428	2	100	0,100408	2	100	0,166591
GistSUMM	4	200	0,097499	3,96	198	0,102399	3	150	0,125469	3,16	158	0,213517

2. Domínio Agronomia

Tabela 44. Teste estatístico da métrica F-Measure – Domínio Agronomia.

Comparando amostras múltiplas relacionadas – F-Measure – Domínio Agronomia												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	135,096	p-nível	0	131,136	p-nível	0	132,696	p-nível	0	144,6	p-nível	0
Coef. de concordância de Kendall	0,90064	Ordem médio	0,898612245	0,87424	Ordem médio	0,871673469	0,88464	Ordem médio	0,882285714	0,964	Ordem médio	0,963265306
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	3	150	0,239569383	3,1	155	0,320931006	3,96	198	0,38170496	4	200	0,440490418
PragmaSUM	1,46	73	0,189194643	1,34	67	0,244484584	1,5	75	0,299103114	2,1	105	0,395708039
PragmaSUM_S em_chaves	1,54	77	0,189073809	1,66	83	0,251456271	1,5	75	0,299883863	1	50	0,364065199
GistSUMM	4	200	0,317268661	3,9	195	0,326505708	3,04	152	0,363986476	2,9	145	0,408345841

Tabela 45. Teste estatístico da métrica Precision – Domínio Agronomia.

Comparando amostras múltiplas relacionadas – Precision – Domínio Agronomia												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	144,168	p-nível	0	145,44	p-nível	0	141,864	p-nível	0	142,056	p-nível	0
Coef. de concordância de Kendall	0,96112	Ordem médio	0,960326531	0,9696	Ordem médio	0,968979592	0,94576	Ordem médio	0,944653061	0,94704	Ordem médio	0,945959184
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	2,04	102	0,852965039	1,98	99	0,753250712	1,96	98	0,661154492	2,06	103	0,49353758
PragmaSUM	3,02	151	0,890006963	3,06	153	0,817295234	3,98	199	0,762912983	4	200	0,643854013
PragmaSUM_S em_chaves	3,94	197	0,895756748	3,94	197	0,825854408	2,98	149	0,749186748	2,9	145	0,593056217
GistSUMM	1	50	0,678060876	1,02	51	0,668927492	1,08	54	0,620380178	1,04	52	0,458266106

Tabela 46. Teste estatístico da métrica Recall – Domínio Agronomia.

Comparando amostras múltiplas relacionadas – Recall – Domínio Agronomia												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	135,216	p-nível	0	134,016	p-nível	0	133,56	p-nível	0	135,096	p-nível	0
Coef. de concordância de Kendall	0,90144	Ordem médio	0,899428571	0,89344	Ordem médio	0,891265306	0,8904	Ordem médio	0,888163265	0,90064	Ordem médio	0,898612245
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	3	150	0,14113709	3,06	153	0,20816893	3,96	198	0,276107053	4	200	0,415366415
PragmaSUM	1,44	72	0,107221203	1,3	65	0,14646069	1,38	69	0,190649393	1,46	73	0,297020027
PragmaSUM_S em_chaves	1,56	78	0,107355412	1,7	85	0,152773392	1,62	81	0,196504697	1,54	77	0,29986047
GistSUMM	4	200	0,218158714	3,94	197	0,226299149	3,04	152	0,266824396	3	150	0,385850447

3. Domínio Educação Física

Tabela 47. Teste estatístico da métrica F-Measure – Domínio Educação Física.

Comparando amostras múltiplas relacionadas – F-Measure – Domínio Educação Física												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	150	p-nível	0	145,584	p-nível	0	148,824	p-nível	0	135,576	p-nível	0
Coef. de concordância de Kendall	1	Ordem médio	1	0,97056	Ordem médio	0,969959184	0,99216	Ordem médio	0,992	0,90384	Ordem médio	0,901877551
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	3	150	0,143723765	3,08	154	0,19682573	4	200	0,243272488	4	200	0,312835093
PragmaSUM	1	50	0,107768603	1	50	0,144701058	1,02	51	0,182048742	2,06	103	0,261346268
PragmaSUM_S em_chaves	2	100	0,11359099	2	100	0,153144104	1,98	99	0,188199644	1,1	55	0,253838604
GistSUMM	4	200	0,192514765	3,92	196	0,205223842	3	150	0,231349413	2,84	142	0,283780899

Tabela 48. Teste estatístico da métrica Precision – Domínio Educação Física.

Comparando amostras múltiplas relacionadas – Precision – Domínio Educação Física												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	135,864	p-nível	0	147,696	p-nível	0	136,176	p-nível	0	134,016	p-nível	0
Coef. de concordância de Kendall	0,90576	Ordem médio	0,903836735	0,98464	Ordem médio	0,984326531	0,90784	Ordem médio	0,905959184	0,89344	Ordem médio	0,891265306
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	2	100	0,86995477	2	100	0,779130681	2	100	0,696931942	1,7	85	0,544743689
PragmaSUM	3,62	181	0,901890805	3,96	198	0,867266399	3,64	182	0,817051212	3,94	197	0,728743377
PragmaSUM_S em_chaves	3,38	169	0,900915281	3,04	152	0,860606193	3,36	168	0,815266664	3,06	153	0,70418346
GistSUMM	1	50	0,737640394	1	50	0,717020338	1	50	0,657694312	1,3	65	0,532596805

Tabela 49. Teste estatístico da métrica Recall – Domínio Educação Física.

Comparando amostras múltiplas relacionadas – Recall – Domínio Educação Física												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	150	p-nível	0	146,616	p-nível	0	150	p-nível	0	137,904	p-nível	0
Coef. de concordância de Kendall	1	Ordem médio	1	0,97744	Ordem médio	0,976979592	1	Ordem médio	1	0,91936	Ordem médio	0,917714286
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	3	150	0,081920474	3,06	153	0,120829226	4	200	0,161610206	4	200	0,253499632
PragmaSUM	1	50	0,059322288	1	50	0,082739582	1	50	0,109019164	1,28	64	0,176128887
PragmaSUM_S em_chaves	2	100	0,063234203	2	100	0,089042565	2	100	0,114689945	1,72	86	0,179933594
GistSUMM	4	200	0,116910757	3,94	197	0,128209413	3	150	0,154059412	3	150	0,222010636

4. Domínio Engenharia

Tabela 50. Teste estatístico da métrica F-Measure – Domínio Engenharia.

Comparando amostras múltiplas relacionadas – F-Measure – Domínio Engenharia												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	146,616	p-nível	0	147,696	p-nível	0	132	p-nível	0	144,6	p-nível	0
Coef. de concordância de Kendall	0,97744	Ordem médio	0,9769795 92	0,98464	Ordem médio	0,9843265 31	0,88	Ordem médio	0,8775510 2	0,964	Ordem médio	0,963265 306
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	3	150	0,1149947 96	3	150	0,1699734 06	3,58	179	0,2214917 48	4	200	0,310613 544
PragmaSUM	1,94	97	0,0915956	1,96	98	0,1327553 17	1,94	97	0,1755099 34	1,9	95	0,267819 689
PragmaSUM_S em_chaves	1,06	53	0,0881810 48	1,04	52	0,1247171 11	1,06	53	0,1642499	1,1	55	0,256610 971
GistSUMM	4	200	0,1987217 41	50	Graus de liberdade	3	3,42	171	0,2200017 58	3	150	0,303667 892

Tabela 51. Teste estatístico da métrica Precision – Domínio Engenharia.

Comparando amostras múltiplas relacionadas – Precision – Domínio Engenharia												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	150	p-nível	0	147,696	p-nível	0	150	p-nível	0	122,4	p-nível	0
Coef. de concordância de Kendall	1	Ordem médio	1	0,98464	Ordem médio	0,984326531	1	Ordem médio	1	0,816	Ordem médio	0,812244898
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	2	100	0,874199199	2	100	0,80616687	2	100	0,731328626	1,38	69	0,593631115
PragmaSUM	3	150	0,900502612	3,04	152	0,871773686	3	150	0,838099713	3,34	167	0,76223706
PragmaSUM_S em_chaves	4	200	0,906384524	3,96	198	0,876865129	4	200	0,843962857	3,66	183	0,764743223
GistSUMM	1	50	0,7082059	1	50	0,703098509	1	50	0,682263903	1,62	81	0,591422376

Tabela 52. Teste estatístico da métrica Recall – Domínio Engenharia.

Comparando amostras múltiplas relacionadas – Recall – Domínio Engenharia												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	146,616	p-nível	0	146,616	p-nível	0	137,76	p-nível	0	144,6	p-nível	0
Coef. de concordância de Kendall	0,97744	Ordem médio	0,9769795 92	0,97744	Ordem médio	0,9769795 92	0,9184	Ordem médio	0,9167346 94	0,964	Ordem médio	0,963265 306
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	3	150	0,0618572 91	3,06	153	0,1208292 26	3,18	159	0,1321744 13	4	200	0,215024 804
PragmaSUM	1,94	97	0,0484412 61	1	50	0,0827395 82	1,94	97	0,0988624 65	1,9	95	0,165137 964
PragmaSUM_S em_chaves	1,06	53	0,0465500 85	2	100	0,0890425 65	1,06	53	0,0919116 78	1,1	55	0,157851 191
GistSUMM	4	200	0,1183943 96	50	Graus de liberdade	3	3,82	191	0,1334286 63	3	150	0,208886 301

5. Domínio Filosofia

Tabela 53. Teste estatístico da métrica F-Measure – Domínio Filosofia.

Comparando amostras múltiplas relacionadas – F-Measure – Domínio Filosofia												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	133,872	p-nível	0	137,976	p-nível	0	139,848	p-nível	0	146,616	p-nível	0
Coef. de concordância de Kendall	0,89248	Ordem médio	0,890285714	0,91984	Ordem médio	0,918204082	0,93232	Ordem médio	0,930938776	0,97744	Ordem médio	0,976979592
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	2,72	136	0,069270305	2,9	145	0,102684687	3,94	197	0,135375559	4	200	0,203279909
PragmaSUM	1	50	0,053180044	1	50	0,072605833	1	50	0,096219813	1,06	53	0,151089269
PragmaSUM_S em_chaves	2,32	116	0,071033694	2,16	108	0,087809212	2,12	106	0,108845119	1,94	97	0,161378064
GistSUMM	3,96	198	0,100031463	3,94	197	0,105858958	2,94	147	0,132567716	3	150	0,195107891

Tabela 54. Teste estatístico da métrica Precision – Domínio Filosofia.

Comparando amostras múltiplas relacionadas – Precision – Domínio Filosofia												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	131,112	p-nível	0	116,136	p-nível	0	113,424	p-nível	0	125,208	p-nível	0
Coef. de concordância de Kendall	0,87408	Ordem médio	0,871510204	0,77424	Ordem médio	0,769632653	0,75616	Ordem médio	0,751183673	0,83472	Ordem médio	0,831346939
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	2,32	116	0,821081769	2	100	0,755420932	1,6	80	0,688728707	1,82	91	0,585907556
PragmaSUM	4	200	0,854088696	4	200	0,819018503	4	200	0,790331848	4	200	0,718048986
PragmaSUM_S em_chaves	2,62	131	0,813101149	2,66	133	0,774438761	2,72	136	0,73398676	2,84	142	0,66340841
GistSUMM	1,06	53	0,748373643	1,34	67	0,744175275	1,68	84	0,691316113	1,34	67	0,584472682

Tabela 55. Teste estatístico da métrica Recall – Domínio Filosofia.

Comparando amostras múltiplas relacionadas – Recall – Domínio Filosofia												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	127,464	p-nível	0	114,648	p-nível	0	121,2	p-nível	0	129,552	p-nível	0
Coef. de concordância de Kendall	0,84976	Ordem médio	0,846693878	0,76432	Ordem médio	0,759510204	0,808	Ordem médio	0,804081633	0,86368	Ordem médio	0,860897959
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	2,22	111	0,036825939	2,64	132	0,056719988	3,8	190	0,078352532	3,88	194	0,131620728
PragmaSUM	1	50	0,027778839	1	50	0,0385658	1	50	0,052307202	1	50	0,087766535
PragmaSUM_S em_chaves	2,98	149	0,052006118	2,62	131	0,061909235	2,4	120	0,075375948	2,28	114	0,11101704
GistSUMM	3,8	190	0,056616207	3,74	187	0,058700494	2,8	140	0,076360454	2,84	142	0,124869474

6. Domínio Fisioterapia

Tabela 56. Teste estatístico da métrica F-Measure – Domínio Fisioterapia.

Comparando amostras múltiplas relacionadas – F-Measure – Domínio Fisioterapia												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	148,824	p-nível	0	147,648	p-nível	0	110,472	p-nível	0	83,784	p-nível	4,72E-18
Coef. de concordância de Kendall	0,99216	Ordem médio	0,992	0,98432	Ordem médio	0,984	0,73648	Ordem médio	0,731102041	0,55856	Ordem médio	0,54955102
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	3	150	0,266005687	3,02	151	0,345598131	3,38	169	0,400105789	2,12	106	0,419623735
PragmaSUM	1,02	51	0,228915237	1,02	51	0,295330718	1,24	62	0,355124403	3,14	157	0,423364133
PragmaSUM_S em_chaves	1,98	99	0,23921905	1,98	99	0,319377274	1,9	95	0,364489854	1,32	66	0,408269111
GistSUMM	4	200	0,367524791	3,98	199	0,369070958	3,48	174	0,400474629	3,42	171	0,433510429

Tabela 57. Teste estatístico da métrica Precision – Domínio Fisioterapia.

Comparando amostras múltiplas relacionadas – Precision – Domínio Fisioterapia												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	145,296	p-nível	0	146,616	p-nível	0	133,56	p-nível	0	146,616	p-nível	0
Coef. de concordância de Kendall	0,96864	Ordem médio	0,968	0,97744	Ordem médio	0,976979592	0,8904	Ordem médio	0,888163265	0,97744	Ordem médio	0,976979592
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	2,04	102	0,845270531	2	100	0,718418891	1,32	66	0,618761302	1	50	0,431787006
PragmaSUM	3,96	198	0,876387319	3,94	197	0,798053449	3,94	197	0,728533671	3,94	197	0,569271618
PragmaSUM_S em_chaves	3	150	0,865277388	3,06	153	0,778070747	3,06	153	0,699522391	3,06	153	0,518927811
GistSUMM	1	50	0,662659501	1	50	0,661755171	1,68	84	0,621751694	2	100	0,45215538

Tabela 58. Teste estatístico da métrica Recall – Domínio Fisioterapia.

Comparando amostras múltiplas relacionadas – Recall – Domínio Fisioterapia												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3	50	Graus de liberdade	3	50	Graus de liberdade	3	50	Graus de liberdade	3
qui-quadrado	148,824	p-nível	0	145,05210 42	p-nível	0	128,472	p-nível	0	125,16	p-nível	0
Coef. de concordância de Kendall	0,99216	Ordem médio	0,992	0,9670140 28	Ordem médio	0,9663408 45	0,85648	Ordem médio	0,8535510 2	0,8344	Ordem médio	0,831020 408
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	3	150	0,1595933 16	2,94	147	0,2313877 59	3,14	157	0,3024220 64	3,22	161	0,422424 914
PragmaSUM	1,02	51	0,1328454 12	1,03	51,5	0,1839919 76	1,06	53	0,2396221 7	1,28	64	0,351554 265
PragmaSUM_S em_chaves	1,98	99	0,1400212 63	2,03	101,5	0,2101672 31	2,04	102	0,2604834 49	1,74	87	0,362775 946
GistSUMM	4	200	0,2627574 29	4	200	0,2641353 74	3,76	188	0,3040518 45	3,76	188	0,433170 035

7. Domínio Geografia

Tabela 59. Teste estatístico da métrica F-Measure – Domínio Geografia.

Comparando amostras múltiplas relacionadas – F-Measure – Domínio Geografia												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	150	p-nível	0	135,024	p-nível	0	132,912	p-nível	0	131,184	p-nível	0
Coef. de concordância de Kendall	1	Ordem médio	1	0,90016	Ordem médio	0,8981224 49	0,88608	Ordem médio	0,8837551 02	0,87456	Ordem médio	0,872
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	3	150	0,1598359 78	3	150	0,2192624 96	3,96	198	0,2682773 9	3,92	196	0,336496 247
PragmaSUM	1	50	0,1361859 6	1,52	76	0,1892209 74	1,56	78	0,2349536 28	1,6	80	0,315025 409
PragmaSUM_S em_chaves	2	100	0,1408149 6	1,48	74	0,1885941 6	1,44	72	0,2345011 25	1,4	70	0,311525 351
GistSUMM	4	200	0,2192957 82	4	200	0,2286839 88	3,04	152	0,264385	3,08	154	0,335387 109

Tabela 60. Teste estatístico da métrica Precision – Domínio Geografia.

Comparando amostras múltiplas relacionadas – Precision – Domínio Geografia												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	143,088	p-nível	0	148,824	p-nível	0	136,944	p-nível	0	144,312	p-nível	0
Coef. de concordância de Kendall	0,95392	Ordem médio	0,9529795 92	0,99216	Ordem médio	0,992	0,91296	Ordem médio	0,9111836 73	0,96208	Ordem médio	0,961306 122
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	2,04	102	0,8601914 95	2,02	101	0,7699285 73	1,68	84	0,6851647 67	1,08	54	0,540026 283
PragmaSUM	3,92	196	0,8944798 61	4	200	0,8510792	4	200	0,7997182 86	4	200	0,699780 687
PragmaSUM_S em_chaves	3,04	152	0,8890212 62	2,98	149	0,8206494 14	3	150	0,7674007 2	2,98	149	0,639533 843
GistSUMM	1	50	0,7324267 2	1	50	0,7242621 96	1,32	66	0,6815477 81	1,94	97	0,553854 912

Tabela 61. Teste estatístico da métrica Recall – Domínio Geografia.

Comparando amostras múltiplas relacionadas – Recall – Domínio Geografia												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	150	p-nível	0	135,096	p-nível	0	133,56	p-nível	0	133,488	p-nível	0
Coef. de concordância de Kendall	1	Ordem médio	1	0,90064	Ordem médio	0,898612245	0,8904	Ordem médio	0,888163265	0,88992	Ordem médio	0,887673469
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	3	150	0,09135585	3	150	0,135675244	3,96	198	0,181525566	3,92	196	0,281561924
PragmaSUM	1	50	0,076033783	1,46	73	0,111627156	1,38	69	0,146994161	1,28	64	0,225894703
PragmaSUM_S em_chaves	2	100	0,078872943	1,54	77	0,112258464	1,62	81	0,148266294	1,72	86	0,231417664
GistSUMM	4	200	0,140710862	4	200	0,147023545	3,04	152	0,17879791	3,08	154	0,277336043

8. Domínio História

Tabela 62. Teste estatístico da métrica F-Measure – Domínio História.

Comparando amostras múltiplas relacionadas – F-Measure – Domínio História												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3	50	Graus de liberdade	3	50	Graus de liberdade	3	50	Graus de liberdade	3
qui-quadrado	150	p-nível	0	150	p-nível	0	150	p-nível	0	140,326 6533	p-nível	0
Coef. de concordância de Kendall	1	Ordem médio	1	1	Ordem médio	1	1	Ordem médio	1	0,93551 1022	Ordem médio	0,934194 92
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	3	150	0,0544485 67	3	150	0,0799925 63	3	150	0,1063675 8	3,04	152	0,169145 909
PragmaSUM	1	50	0,0459094 72	1	50	0,0656298 55	1	50	0,0878820 79	1,85	92,5	0,143783 079
PragmaSUM_S em_chaves	2	100	0,0477927 3	2	100	0,0676922 77	2	100	0,0906718 38	1,15	57,5	0,141792 715
GistSUMM	4	200	0,0772380 54	4	200	0,0854434 6	4	200	0,1097995 91	3,96	198	0,174580 64

Tabela 63. Teste estatístico da métrica Precision – Domínio História.

Comparando amostras múltiplas relacionadas – Precision – Domínio História												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	143,304	p-nível	0	135,024	p-nível	0	150	p-nível	0	146,52	p-nível	0
Coef. de concordância de Kendall	0,95536	Ordem médio	0,95444898	0,90016	Ordem médio	0,898122449	1	Ordem médio	1	0,9768	Ordem médio	0,976326531
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	2,02	101	0,837017473	1,52	76	0,756246618	1	50	0,695847958	1,04	52	0,595269919
PragmaSUM	3,9	195	0,860531931	4	200	0,826881066	4	200	0,791909498	3,98	199	0,725108908
PragmaSUM_S em_chaves	3,08	154	0,858170145	3	150	0,818270934	3	150	0,780893418	3,02	151	0,701708984
GistSUMM	1	50	0,771430768	1,48	74	0,76057084	2	100	0,718453552	1,96	98	0,624610429

Tabela 64. Teste estatístico da métrica Recall – Domínio História.

Comparando amostras múltiplas relacionadas – Recall – Domínio História												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	150	p-nível	0	150	p-nível	0	150	p-nível	0	136,752	p-nível	0
Coef. de concordância de Kendall	1	Ordem médio	1	1	Ordem médio	1	1	Ordem médio	1	0,91168	Ordem médio	0,909877551
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	3	150	0,028393231	3	150	0,042844247	3	150	0,058752402	3,04	152	0,102197552
PragmaSUM	1	50	0,023765868	1	50	0,034581009	1	50	0,047321213	1,76	88	0,082192457
PragmaSUM_S em_chaves	2	100	0,024768547	2	100	0,035732157	2	100	0,049059034	1,24	62	0,08158748
GistSUMM	4	200	0,04109046	4	200	0,045873703	4	200	0,060675218	3,96	198	0,105260868

9. Domínio Medicina

Tabela 65. Teste estatístico da métrica F-Measure – Domínio Medicina.

Comparando amostras múltiplas relacionadas – F-Measure – Domínio Medicina												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	147,696	p-nível	0	146,472	p-nível	0	132,56	p-nível	0	103,56	p-nível	0
Coef. de concordância de Kendall	0,98464	Ordem médio	0,984326531	0,97648	Ordem médio	0,976	0,883733333	Ordem médio	0,881360544	0,6904	Ordem médio	0,684081633
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	3	150	0,220498236	2,98	149	0,292969797	3,74	187	0,343857726	2,24	112	0,374678358
PragmaSUM	1,04	52	0,19186213	1,04	52	0,25553467	1,52	76	0,306347272	3,06	153	0,380745888
PragmaSUM_S em_chaves	1,96	98	0,197186397	1,98	99	0,261816534	1,52	76	0,306347272	1,12	56	0,358818245
GistSUMM	4	200	0,272439115	4	200	0,298991201	3,22	161	0,340003247	3,58	179	0,382185309

Tabela 66. Teste estatístico da métrica Precision – Domínio Medicina.

Comparando amostras múltiplas relacionadas – Precision – Domínio Medicina												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	149,4216867	p-nível	0	146,52	p-nível	0	147,44	p-nível	0	138,456	p-nível	0
Coef. de concordância de Kendall	0,996144578	Ordem médio	0,996065896	0,9768	Ordem médio	0,976326531	0,982933333	Ordem médio	0,982585034	0,92304	Ordem médio	0,921469388
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	2	100	0,817999772	1,98	99	0,718002185	1,96	98	0,629952171	1,26	63	0,45136252
PragmaSUM	3,98	199	0,86236162	3,96	198	0,802144932	3,5	175	0,725954911	4	200	0,592140146
PragmaSUM_S em_chaves	3,02	151	0,852849145	3,04	152	0,783069248	3,5	175	0,725954911	3	150	0,538542722
GistSUMM	1	50	0,712659289	1,02	51	0,679190342	1,04	52	0,618110554	1,74	87	0,462147432

Tabela 67. Teste estatístico da métrica Recall – Domínio Medicina.

Comparando amostras múltiplas relacionadas – Recall – Domínio Medicina												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3	50	Graus de liberdade	3	50	Graus de liberdade	3	50	Graus de liberdade	3
qui-quadrado	147,696	p-nível	0	147,696	p-nível	0	137,173 3333	p-nível	0	115,848	p-nível	0
Coef. de concordância de Kendall	0,98464	Ordem médio	0,9843265 31	0,98464	Ordem médio	0,9843265 31	0,91448 8889	Ordem médio	0,9127437 64	0,77232	Ordem médio	0,767673 469
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	3	150	0,1313799 39	3	150	0,1924919 62	3,74	187	0,2518763 44	3,48	174	0,360074 468
PragmaSUM	1,04	52	0,1105922 35	1,04	52	0,1577660 81	1,5	75	0,2044933 38	1,62	81	0,304847 691
PragmaSUM_S em_chaves	1,96	98	0,1143700 33	1,96	98	0,1637465 74	1,5	75	0,2044933 38	1,42	71	0,296487 819
GistSUMM	4	200	0,1786293 24	4	200	0,2024250 33	3,26	163	0,2494745 07	3,48	174	0,359503 931

10. Domínio Psicologia

Tabela 68. Teste estatístico da métrica F-Measure – Domínio Psicologia.

Comparando amostras múltiplas relacionadas – F-Measure – Domínio Psicologia												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	150	p-nível	0	150	p-nível	0	141,072	p-nível	0	136,728	p-nível	0
Coef. de concordância de Kendall	1	Ordem médio	1	1	Ordem médio	1	0,94048	Ordem médio	0,939265306	0,91152	Ordem médio	0,909714286
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	3	150	0,079642967	3	150	0,115519786	2,92	146	0,153588368	3,18	159	0,234158374
PragmaSUM	1	50	0,06521416	1	50	0,09234146	1	50	0,122861559	1,08	54	0,197478962
PragmaSUM_S em_chaves	2	100	0,070054176	2	100	0,101254291	2,12	106	0,139615444	1,92	96	0,217250278
GistSUMM	4	200	0,122106643	4	200	0,134011763	3,96	198	0,158382498	3,82	191	0,23646235

Tabela 69. Teste estatístico da métrica Precision – Domínio Psicologia.

Comparando amostras múltiplas relacionadas – Precision – Domínio Psicologia												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	149,7054108	p-nível	0	148,824	p-nível	0	136,176	p-nível	0	143,304	p-nível	0
Coef. de concordância de Kendall	0,998036072	Ordem médio	0,997995992	0,99216	Ordem médio	0,992	0,90784	Ordem médio	0,905959184	0,95536	Ordem médio	0,95444898
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	2	100	0,851287083	2,02	101	0,762369894	1,36	68	0,692953376	1,1	55	0,569931999
PragmaSUM	3,99	199,5	0,885203963	4	200	0,847928852	4	200	0,806515542	4	200	0,733259772
PragmaSUM_S em_chaves	3,01	150,5	0,877538534	2,98	149	0,816117608	3	150	0,754941574	2,98	149	0,636645448
GistSUMM	1	50	0,739816334	1	50	0,722135901	1,64	82	0,689676012	1,92	96	0,599411154

Tabela 70. Teste estatístico da métrica Recall – Domínio Psicologia.

Comparando amostras múltiplas relacionadas – Recall – Domínio Psicologia												
	50%			70%			80%			90%		
N	50	Graus de liberdade	3									
qui-quadrado	150	p-nível	0	150	p-nível	0	128,904	p-nível	0	126,024	p-nível	0
Coef. de concordância de Kendall	1	Ordem médio	1	1	Ordem médio	1	0,85936	Ordem médio	0,856489796	0,84016	Ordem médio	0,836897959
	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média	Ordem médio	Soma de ordens	Média
BLMSumm	3	150	0,041943382	3	150	0,062845323	2,82	141	0,087083775	3,3	165	0,149992006
PragmaSUM	1	50	0,033941955	1	50	0,04905004	1	50	0,067012959	1	50	0,115929623
PragmaSUM_S em_chaves	2	100	0,03658626	2	100	0,054375115	2,3	115	0,081885704	2,12	106	0,140590419
GistSUMM	4	200	0,067552974	4	200	0,074626353	3,88	194	0,090353684	3,58	179	0,149742224

ANEXO

