* **SemanticSplitterNodeParser (splitter)**

|  |  |  |  |
| --- | --- | --- | --- |
| Parametreler | text1.txt (331 kelime) | text2.txt (478 kelime, iki ayrı hikaye) | text3.txt (846 kelime, 4 ayrı konu) |
| buffer\_size=1, breakpoint\_percentile\_threshold=95, embed\_model=embed\_model | 3 chunk, birbiriyle uyumlu bir şekilde bölmüş | 4 chunk, 4. chunkta iki ayrı hikayeyi birleştirmiş, 3. chunk tek cümlelik oldu | 5 chunk, 2.chunkta 3 ayrı konunun parçalarını almış |
| buffer\_size=1, breakpoint\_percentile\_threshold=90, embed\_model=embed\_model | 4 chunk, birbiriyle uyumlu bir şekilde bölmemiş, bir chunkın devamı olacak cümle diğer chunkın başına eklenmiş | 7 chunk, 5. chunkta chunkta iki ayrı hikayeyi birleştirmiş, | 9 chunk, 2. chunkta ilk iki yazıyı birleştirmiş |
| buffer\_size=1, breakpoint\_percentile\_threshold=100, embed\_model=embed\_model | 1 chunk | 1 chunk | 1 chunk |
| buffer\_size=2, breakpoint\_percentile\_threshold=95, embed\_model=embed\_model | 3 chunk, birbiriyle uyumlu bir şekilde bölmemiş, bir chunın devamı olacak cümle diğer chunkın başına eklenmiş | 4 chunk, 2. chunkta iki ayrı hikayeyi birleştirmiş, 3. chunk tek cümlelik oldu | 5 chunk, 2. chunkta 2. yazının tamamını 3. yazının bir parçasını birleştirmiş. |
| buffer\_size=2, breakpoint\_percentile\_threshold=90, embed\_model=embed\_model | 4 chunk, birbiriyle uyumlu bir şekilde bölmemiş, bir chunkın devamı olacak cümle diğer chunkın başına eklenmiş | 7 chunk, 3. chunkta chunkta iki ayrı hikayeyi birleştirmiş, | 9 chunk, biraz daha iyi parçalamış. |
| buffer\_size=2, breakpoint\_percentile\_threshold=100, embed\_model=embed\_model | 1 chunk | 1 chunk | 1 chunk |

Text dısyaları için yazılar chatgpt’ye yazdırdım.

Bazı yazıları tam olarak bitirmeden, örneğin son cümlesini diğer parçanın başına eklemiş. Ama parçalarda anlam uyumu var genel olarak.

* **SentenceSplitter (base\_splitter)**

|  |  |  |  |
| --- | --- | --- | --- |
| parametreler | text1.txt (331 kelime) | text2.txt (478 kelime, iki ayrı hikaye) | text3.txt (846 kelime, 4 ayrı konu) |
| chunk\_size=512, chunk\_overlap = 0 | 2 chunk, 2. chunkın ilk cümlesini 1.chunkta bırakmış, başarılı bölmemiş | 4 chunk, 1. chunkta iki hikayeyi birleştirmiş | 6 chunk, yazıları dengeli parçalamamış |
| chunk\_size=300, chunk\_overlap = 0 | 3 chunk, başarılı bölmemiş | 6 chunk, 3. chunkta ikinci hikayeye başlamış | 10 chunk, yazıları dengeli parçalamamış |
| chunk\_size=512, chunk\_overlap = 20 | 2 chunk, 2. chunkın ilk cümlesini 1.chunkta bırakmış, başarılı bölmemiş | 4 chunk, 1. chunkta iki hikayeyi birleştirmiş | 6 chunk, yazıları dengeli parçalamamış |
| chunk\_size=300, chunk\_overlap = 20 | 3 chunk, aynı cümle 1. ve 2. chunkta yer alır, başarılı bölmemiş | 6 chunk, 3. chunkta iki hikayeyi birleştirmiş | 10 chunk, 4. chunkta 2. yazı ile 3. yazıyı birleştirmiş |