

Domain	Method	Input	Noise (↓)	Complete (↑)	Context Switch (↓)	Task Relevance (↑)
Support	Recursive	Text	27.56	15.36	23.60	84.38
	Recursive	HTML	25.59	55.75	2.96	45.16
	Recursive	Markdown	26.46	27.34	24.34	82.32
	Embedding	Markdown	35.86	9.41	59.41	57.92
	LLMSemantic	Markdown	24.00	71.21	6.81	76.89
	LumberChunker	Markdown	36.05	1.25	54.64	63.16
	AutoChunker	Markdown	<b>1.12</b>	<b>93.03</b>	<b>1.66</b>	<b>94.76</b>
	Recursive	Text	29.83	18.45	25.12	82.54
	Recursive	HTML	26.91	53.62	3.15	47.23
Wikipedia	Recursive	Markdown	28.13	25.67	26.45	80.91
	Embedding	Markdown	37.42	8.92	61.23	55.84
	LLMSemantic	Markdown	25.34	69.87	7.12	75.32
	LumberChunker	Markdown	38.21	2.14	56.78	61.45
	AutoChunker	Markdown	<b>2.31</b>	<b>91.24</b>	<b>2.05</b>	<b>92.87</b>

Table 2: Comparison of Different Chunking Techniques Across Domains. ↑ indicates higher is better, ↓ indicates lower is better. Best results are in **bold**.

Domain	Method	WP@1	WP@3	WP@5
Support	Recursive	60.75	51.25	39.15
	Embedding	16.75	14.25	13.65
	LLMSemantic	69.12	56.23	49.41
	AutoChunker	<b>75.42</b>	<b>63.42</b>	<b>56.84</b>
	AutoChunker + CAR	<b>75.42</b>	<b>68.74</b>	<b>63.22</b>
Wikipedia	Recursive	58.45	48.92	37.84
	Embedding	15.92	13.85	12.95
	LLMSemantic	66.78	54.32	47.65
	AutoChunker	<b>72.95</b>	<b>61.45</b>	<b>54.92</b>
	AutoChunker + CAR	<b>72.95</b>	<b>66.84</b>	<b>61.35</b>

Table 3: Comparison of Weighted Precision Scores Across Different Methods and Domains. CAR: Context Aware Retrieval. Best results are in **bold**.

- **Embedding:** We converted HTML content to markdown and utilized Langchain’s SemanticChunker (Chase, 2022) with *cohere.embed-multilingual-v3* (Cohere, 2023).
- **LLMSemantic:** We used the code provided by the authors, employing the *claude-3.5-sonnet* (Anthropic, 2024) model as the LLM backbone.
- **LumberChunker:** We implemented this method using the code provided by the authors, also using the *claude-3.5-sonnet* model as the LLM backbone.

We used *claude-3.5-sonnet* for AutoChunker and all LLM-based evaluations, and *cohere.embed-multilingual-v3* as the embedding model for the retriever.

## 6 Results and Analysis

### 6.1 Chunking Quality Analysis

Table 2 presents the results comparing different chunking techniques across various metrics. Our approach significantly outperforms all baselines across all metrics. It achieves the lowest noise, highest completeness, minimal context switching, and highest task relevancy. The substantial reduction in noise can be attributed to our elimination mechanism, which addresses a critical gap in existing techniques.

### 6.2 Retrieval Performance

We evaluated the retrieval performance using weighted precision scores at different ranks. Table 3 shows these results. Our method consistently outperforms baselines in retrieval performance, with the highest WP@1. The addition of information via Context Aware Retrieval (CAR) further improves