

Expansão da Ferramenta Doccano para a Anotação de Análise de Sentimento Estruturada

Antonio Paulo Martins Ferri¹, Guilherme Dallmann Lima¹,
Larissa Astrogildo de Freitas¹, Andrew Campos¹

¹Universidade Federal de Pelotas (UFPel)
Pelotas – RS – Brasil

apmferri@inf.ufpel.edu.br, gdlima@inf.ufpel.edu.br
larissa@inf.ufpel.edu.br, abdcampos@inf.ufpel.edu.br

Abstract. *Annotation tools are essential for building datasets in Natural Language Processing (NLP). However, the Doccano version adapted for Aspect-Based Sentiment Analysis (ABSA) does not explicitly represent the author of an opinion. This work presents an extension that incorporates the opinion holder into the annotation structure, modifying the data model, import/export workflows, and annotation interface. A preliminary evaluation annotated 30 hotel reviews in both versions: 10 contained an explicitly identifiable opinion holder, cases in which the extension enabled a representation unavailable in the baseline version.*

Resumo. *Ferramentas de anotação são fundamentais para a construção de conjuntos de dados em Processamento de Linguagem Natural (PLN). Contudo, a versão do Doccano adaptada para Análise de Sentimento Baseada em Aspectos (ASBA) não representa explicitamente o autor de uma opinião. Este trabalho apresenta uma extensão dessa ferramenta que incorpora o opinion holder à estrutura de anotação, modificando o modelo de dados, os fluxos de importação e exportação e a interface. A avaliação preliminar anotou 30 avaliações de hotéis em ambas as versões: 10 continham um autor explicitamente identificável, casos em que a extensão permitiu registrar uma associação indisponível na versão base.*

1. Introdução

A construção de conjuntos de dados anotados constitui uma etapa importante no desenvolvimento e avaliação de modelos de Processamento de Linguagem Natural (PLN). Nesse contexto, ferramentas como o Doccano [Nakayama et al. 2018] auxiliam na anotação de textos para diferentes tarefas, incluindo análise de sentimentos.

Nesse cenário, a Análise de Sentimento Baseada em Aspectos (ASBA) busca identificar aspectos mencionados em um texto e os sentimentos associados a eles. Tarefas de ASBA frequentemente envolvem subtarefas como Extração de Aspecto e Classificação de Sentimento do Aspecto [Pontiki et al. 2016]. Embora a literatura de mineração de opiniões reconheça o autor da opinião (*opinion holder*) como um elemento relevante da estrutura opinativa [Liu 2012], sua representação explícita não está necessariamente presente em esquemas e ferramentas voltados à ASBA. A Análise de Sentimento Estruturada (ASE), por sua vez, inclui o emissor da opinião e sua expressão de sentimento [Barnes et al. 2021], ampliando a informação a respeito da opinião.

O objetivo deste trabalho é expandir a versão do Doccano previamente adaptada para ASBA [Cardozo 2025], incorporando a representação explícita do autor da opinião. Para isso, foi introduzido o *span* “autor” e sua associação às relações entre aspecto e opinião. As principais contribuições do trabalho são: (i) extensão do modelo de dados

para representar autoria; (ii) adaptação dos fluxos de importação e exportação; e (iii) modificação da interface para criação e visualização dessas relações.

A representação explícita do autor permite distinguir situações em que diferentes indivíduos expressam opiniões sobre um mesmo aspecto [Wiebe et al. 2005]. Dessa forma, a estrutura anotada passa a preservar não apenas o conteúdo opinativo e sua polaridade, mas também a associação entre a opinião e seu respectivo emissor.

2. Metodologia

2.1. Limitações da Extensão do Doccano para ASBA

A extensão do Doccano para ASBA permite a marcação de aspectos e opiniões, representados, respectivamente, pelos *spans* “aspecto” e “opinião”, relacionados pelas polaridades “positiva”, “negativa” e “neutra”. Entretanto, essa estrutura não representa explicitamente o autor responsável por cada opinião. Em contextos com múltiplos indivíduos, essa ausência pode gerar ambiguidades na associação entre opiniões e seus respectivos autores [Wiebe et al. 2005]. A estrutura de anotação disponível para ASBA é ilustrada na Figura 1.

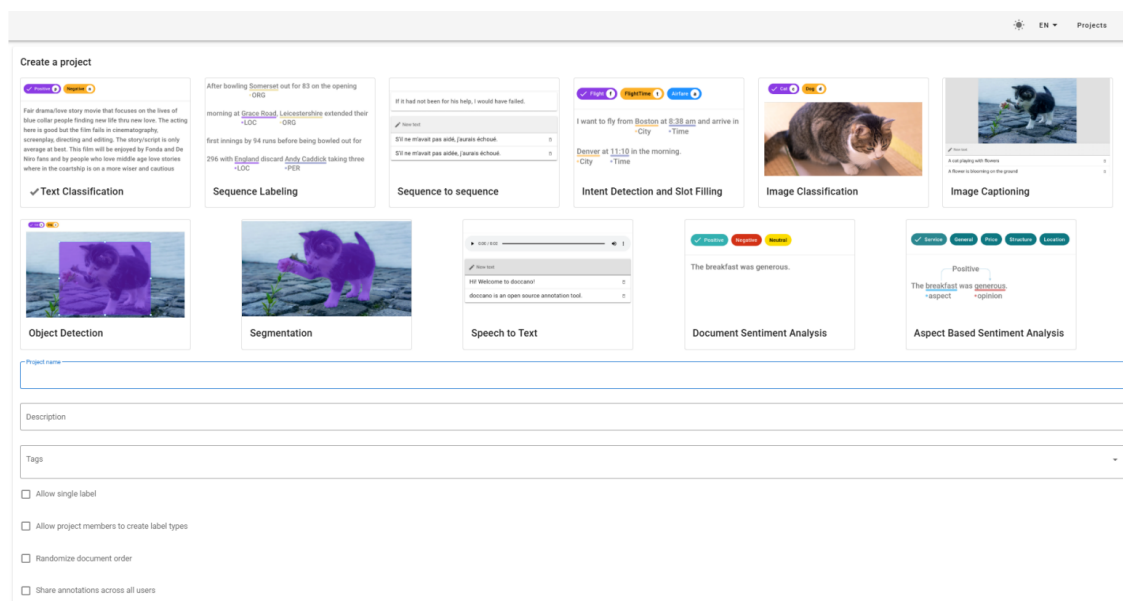


Figura 1. Anotação na ferramenta Doccano para a tarefa de ASBA.

2.2. Implementação

A implementação foi realizada sobre a versão do Doccano previamente expandida para ASBA [Cardozo 2025]. A extensão proposta incorpora o autor da opinião (*opinion holder*) à estrutura já existente, permitindo representar anotações compostas por aspecto, opinião, polaridade e autor.

A decisão de incorporar o autor como uma referência opcional associada à relação existente teve como objetivo preservar a estrutura já utilizada pela versão para ASBA, evitando a criação de um fluxo de anotação paralelo. Dessa forma, a extensão mantém a associação entre aspecto, opinião e polaridade e adiciona a autoria como informação complementar da mesma estrutura relacional.

Tabela 1. Principais alterações realizadas na extensão do Doccano.

Componente	Alteração
Modelo de dados	Inclusão do autor nas relações e persistência no banco de dados.
Importação	Leitura do autor e de sua posição no texto.
Exportação	Inclusão do autor nas anotações exportadas.
Frontend	Criação, associação e visualização da nova relação.

No *backend*, o modelo de relações foi estendido com uma referência opcional ao *span* que representa o autor, incluindo sua persistência no banco de dados. Os fluxos de importação, serialização e exportação também foram adaptados para processar essa informação. Em particular, a importação passou a reconhecer o autor e suas posições no texto, criar os tipos de *span* e relação necessários e persistir as anotações estruturadas. Também foi incluído um teste automatizado para verificar a preservação da informação de autoria no processo de exportação das anotações.

No *frontend*, a representação das relações e o fluxo de anotação foram modificados para incorporar o autor. Quando o anotador relaciona um autor a um aspecto, a interface não cria uma relação independente; em vez disso, atualiza a informação de autoria da relação aspecto–opinião correspondente e utiliza uma ligação auxiliar para representar visualmente essa associação. Essa estratégia preserva o fluxo original de anotação e evita a duplicação de relações no modelo de dados, conforme ilustrado na Figura 2.

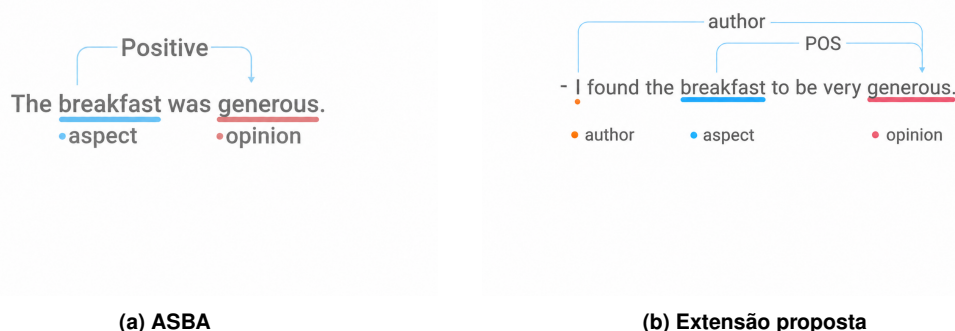


Figura 2. Comparação entre a anotação ASBA (a) e a extensão proposta com representação explícita do autor da opinião (b).

2.3. Avaliação Preliminar

A avaliação preliminar teve como objetivo verificar se a extensão proposta permite representar explicitamente informações de autoria em situações nas quais a estrutura utilizada na versão para ASBA não oferece esse recurso. Para isso, foi realizado um estudo de caso com 30 avaliações de hotéis extraídas de um conjunto de dados previamente utilizado em tarefas de análise de sentimentos estruturada [Campos et al. 2026].

Os mesmos exemplos foram anotados por um único anotador nas duas versões da ferramenta. Em cada caso, foram mantidos os mesmos elementos de aspecto,

opinião e polaridade, permitindo observar especificamente a diferença introduzida pela representação do autor da opinião. Considerou-se como autor uma menção explicitamente identificável no texto à qual pudesse ser atribuída a opinião anotada.

Por exemplo, em um caso no qual uma pessoa é mencionada explicitamente como responsável por determinada opinião, a versão para ASBA representa apenas aspecto, opinião e polaridade, enquanto a extensão também registra a entidade responsável e sua associação à opinião.

Das 30 avaliações analisadas, 10 continham referência explícita a um autor da opinião. Nesses casos, apenas a extensão proposta permitiu representar e armazenar explicitamente a associação entre autor e opinião. O estudo permitiu verificar o funcionamento do novo esquema de anotação, embora uma avaliação com múltiplos anotadores ainda seja necessária para investigar aspectos como usabilidade e concordância interanotador.

3. Conclusão

Este trabalho apresentou uma extensão da versão do Doccano voltada à ASBA, incorporando a representação explícita do autor da opinião. A proposta adiciona esse elemento ao modelo de relações e adapta os fluxos de importação, exportação e anotação da ferramenta, permitindo representar aspecto, opinião, polaridade e o respectivo autor. A avaliação preliminar com avaliações de hotéis permitiu verificar o funcionamento dessa estrutura em comparação à versão utilizada como base. Das 30 avaliações analisadas, 10 continham um autor explicitamente identificável, casos em que a extensão permitiu armazenar a associação entre autor e opinião.

Como limitação, a avaliação foi conduzida por um único anotador, não sendo possível concluir sobre a usabilidade da ferramenta ou a concordância entre diferentes usuários. Como trabalhos futuros, pretende-se realizar experimentos com múltiplos anotadores, avaliar a concordância interanotador e investigar o impacto da informação de autoria em tarefas posteriores de análise de sentimento.

Referências

- Barnes, J., Kurtz, R., Oepen, S., Øvrelid, L., and Velldal, E. (2021). Structured sentiment analysis as dependency graph parsing. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3387–3402.
- Campos, A. B., Corrêa, U. B., and Freitas, L. A. (2026). Structured sentiment analysis in brazilian portuguese: An exploratory study using bertimbau. In *Proceedings of the International Conference on Computational Processing of Portuguese (PROPOR)*. In Portuguese.
- Cardozo, A. d. S. (2025). doccano: Ferramenta de anotação para análise de sentimentos expandida. <https://github.com/aleciocardozo/doccano>. Repositório GitHub.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.

- Nakayama, H., Kubo, T., Kamura, J., Taniguchi, Y., and Liang, X. (2018). doccano: Text annotation tool for human.
- Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., et al. (2016). Semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, pages 19–30.
- Wiebe, J., Wilson, T., and Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, 39:165–210.