Spinfo Cologne at SemEval-2026 Task 4: Explainable Creation of Narrativity Embeddings

Janis Pagel and **Nils Reiter**
Department of Digital Humanities
University of Cologne
{janis.pagel|nils.reiter}@uni-koeln.de

Abstract

We describe our submission to SemEval-2026 Task 4: Narrative Story Similarity and Narrative Representation Learning. The task requires (i) selecting, for a given anchor story summary, which of two candidate summaries is narratively closer (Track A) and (ii) producing a narrative representation of a story as a vector embedding (Track B). Our approach emphasizes interpretability by explicitly eliciting three narrativity aspects with a prompted large language model. We then construct a fixed-size narrative embedding by concatenating aspect-wise representations, comparing a static-embedding baseline (GloVe) to contextualized sentence-transformer embeddings (all-MiniLM-L6-v2). On the development set, the sentence-transformer variant outperforms the static baseline and achieves 61.5% accuracy on Track A, while the GloVe variant performs near chance. Our official submission reaches 60.25% accuracy on the Track A test set and 57.75% accuracy on Track B. Additional ablations show that the aspect pipeline slightly outperforms raw-text embeddings, but that aspect contributions are uneven. Qualitative analysis suggests that failures often stem from inconsistent aspect generation and from overemphasizing theme overlap over event-level similarity.

1 Introduction

Narrativity, at least in one reading of the term, describes the way a story is told. While historically ambiguous and difficult to define, a core aspect of narrativity appears to be the sequential and intersequential telling of plot points in a narrative (Abbott, 2014).

Task 4: "Narrative Story Similarity and Narrative Representation Learning" at SemEval-2026 (Hatzel et al., 2026) covers two tracks of the computational processing of narrativity: (a) the classification of which two texts are narratively more similar (or dissimilar) than other texts and (b) how to represent stories in such a way that narrative similarity can be estimated as cosine similarity. In this task, narrative similarity is defined via three aspects: (i) the main topic or theme of a story, (ii) central events and plot points of the story, and (iii) the outcome of the story, and the core of the task is to produce a vector given a story (snippet).

We approach the task with the aim of balancing predictive performance with interpretability. Instead of directly embedding full stories with a end-to-end encoder, we first elicit a structured representation of narrativity via a prompted LLM and then derive narrative embeddings from these explicit narrativity dimensions. This yields intermediate textual snippets (themes, plot points, outcomes) that can be inspected and qualitatively compared across texts, while still enabling vector-based similarity computations.

Our submission reflects two design assumptions. First, the three aspects provided by the task definition can be operationalized individually by prompting an LLM to produce constrained, specific textual descriptions. Second, narrative similarity can be approximated by comparing concatenated embeddings of these aspect descriptions rather than embeddings of the raw story text. We additionally test this assumption with a raw-text baseline and with aspect ablations. We evaluate two vectorization strategies for converting aspect descriptions into narrative embeddings and report results on the official development and test sets for both tracks.

2 Related Work

Narrative data and story summaries: The shared task builds on research that uses story summaries as proxies for narrative content, enabling large-scale modeling and evaluation. Hatzel and Biemann (2024b) introduce a large dataset of multiple summaries per story where narratively similar texts should map close together while avoiding su-

Proceedings of the The 20th International Workshop on Semantic Evaluation (2026), pages 2376–2383

perficial overlap such as named entities.

Narrative embeddings: Several approaches aim to represent narratives beyond surface semantics. Srivastava and Jojic (2018) propose a probabilistic model that embeds text segments into a spatial structure capturing long-range discourse and sequentiality. Wilner et al. (2021) model narrative event representations via attention-based recontextualization. Within the recent "story embedding" line, Hatzel and Biemann (2023) explore narrative chain embeddings trained with a narrative cloze objective and discuss challenges for narratively grounded similarity, while Hatzel and Biemann (2024a) introduce story embeddings as narrative-focused representations and evaluate them on retrieval and narrative understanding tasks. (Sterner et al., 2026) use contrastive learning with 'narrative twins' (variants) of a narrative to infer story embeddings, which in turn are used to determine salience of events with several strategies.

Text similarity and interpretability: Narrative similarity is closely related to semantic similarity, but requires sensitivity to discourse structure and plot progression. Dimension-based similarity models offer one path towards interpretability by exposing which aspects contribute to similarity judgments (Hatzel et al., 2023). A broader perspective is given by Opitz et al. (2025), who survey methods for interpretable embeddings and similarity explanations, highlighting trade-offs between transparency and performance.

3 System Description

We tackle both tracks with a unified pipeline (Figure 1). Given a story text, we first elicit a structured description of the narrative aligned with the task definition and then map this description into a fixed-size narrative embedding. For track A, narrative similarity between two texts is computed via cosine similarity of their narrative embeddings; for track B, the narrative embedding itself is the desired output.

Step 1: Aspect elicitation via prompting: We prompt a generative LLM to produce aspect-specific textual outputs for (i) abstract themes, (ii) course of action, and (iii) outcomes, following the shared-task annotation guidelines (Hatzel et al., 2025). To encourage consistent structure and reduce verbosity, we enforce output constraints: up to five one-word themes, up to five plot points of at most ten words each, and one outcome of at most ten words. This step aims to disentangle narrativity aspects and to externalize them in a human-readable form, which supports interpretability while still enabling downstream embedding.

Step 2: Constructing narrative embeddings: We convert the elicited aspect texts into a single vector representation per story using one of two methods.

(i) Static word embeddings with aspect averaging: Each token of the elicited sentences is mapped to a static embedding vector. For each narrativity aspect, we compute an "aspect vector" by averaging its token vectors. Finally, we concatenate the aspect vectors (themes, course of action, outcomes) into a single narrative embedding.

(ii) Sentence-transformer embeddings per aspect: For each aspect text, we compute a sentence embedding using a sentence transformer and take the pooler output (Reimers and Gurevych, 2019). We then concatenate the aspect embeddings into a single narrative embedding.

Both methods yield one embedding per text, intended to emphasize narrative content expressed via the three aspects rather than surface overlap in the original story text.

Step 3: Similarity and representation output: For track A, we compute cosine similarity between the embeddings of the anchor text and each candidate text and select the candidate with higher similarity. For track B, we output the computed embedding per text as the narrative representation.

4 Experimental Setup

Data: We develop our system on the development set of the shared task, consisting of 200 triples (anchor, text A, text B) with a label indicating whether text A or text B is narratively closer to the anchor text.[1] The texts are Wikipedia summaries of literary works and movies (Hatzel and Biemann, 2024b).

Aspect prompting: For each story, we generate aspect texts for themes, course of action, and outcomes using a single prompt template (Appendix A, Figure 2) that contains task guideline excerpts and the aforementioned explicit output constraints. The goal is to reduce prompt drift and keep aspect

[1] https://narrative-similarity-task.github.io/data/

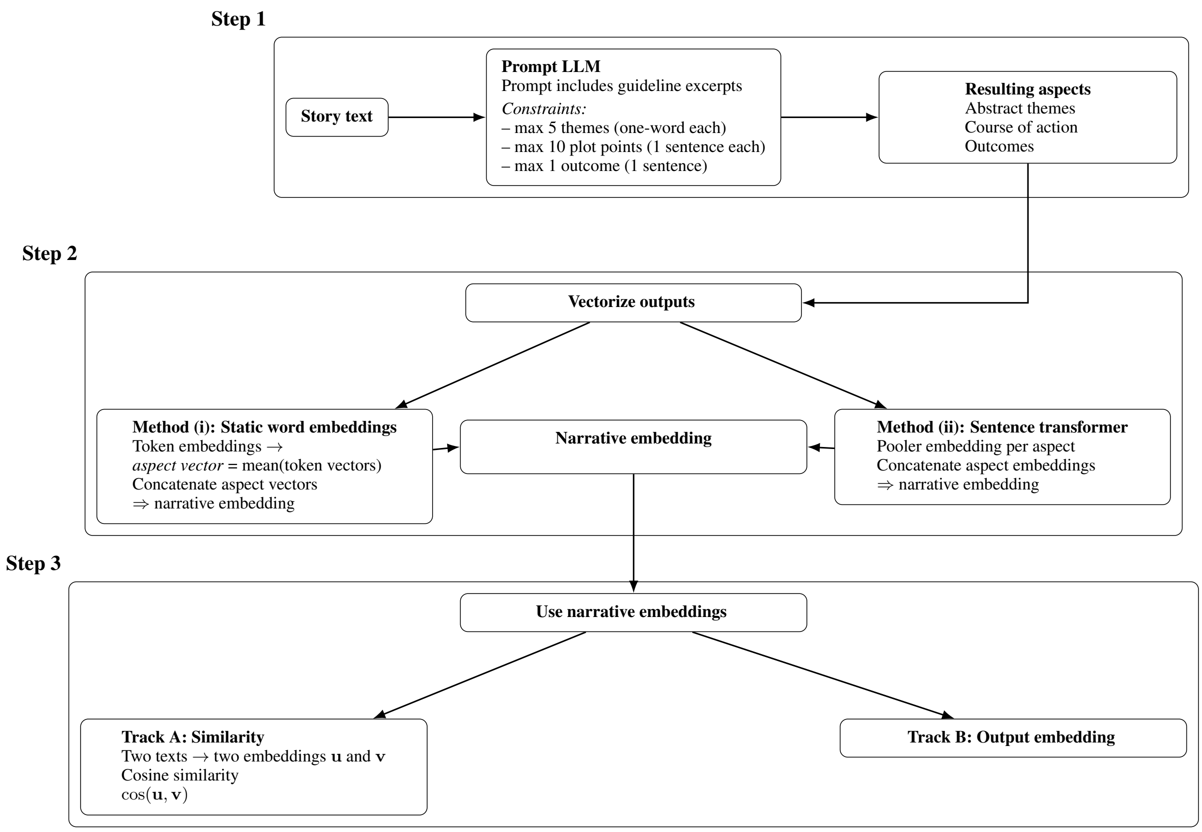

Figure 1: Schematic overview of the system.

outputs comparable across stories. The prompted model is `openai/gpt-4o-mini`.

Embedding configurations: We evaluate two vectorization strategies (Step 2).

(i) Static word embeddings: We use English GloVe embeddings (Pennington et al., 2014) of 300 dimensions and trained on a 6B token corpus.[2] For each aspect, we average token embeddings; the final narrative embedding is the concatenation of all aspect vectors. This setup yields a transparent aggregation mechanism in which each dimension is directly inherited from the underlying static embedding space.

(ii) Sentence-transformer embeddings: We use `all-MiniLM-L6-v2` to embed each aspect text and concatenate aspect embeddings.[3] This setup provides contextualized aspect representations while retaining interpretability at the aspect level.

Similarity for track A: Given an anchor embedding $\mathbf{anchor}$ and candidate embeddings $\mathbf{a}$ and $\mathbf{b}$, we compute cosine similarities $\cos(\mathbf{anchor}, \mathbf{a})$ and $\cos(\mathbf{anchor}, \mathbf{b})$ and predict the candidate with higher similarity to be narratively closer to the anchor.

Ablation studies: We run three additional analyses on the development triples. First, we compare against a raw-text baseline that directly embeds the full anchor and candidate summaries with `all-MiniLM-L6-v2`, without aspect elicitation. Second, we perform aspect ablations using only themes, only plot points, only outcomes, and by dropping one aspect at a time. Third, we test simple weighted combinations of the three aspect-wise similarities. These weights are not learned; instead, we evaluate a small manually defined grid that emphasizes one aspect at a time: $(0.6, 0.2, 0.2)$, $(0.2, 0.6, 0.2)$, $(0.2, 0.2, 0.6)$, $(0.1, 0.7, 0.2)$, and $(0.1, 0.2, 0.7)$, where the values correspond to themes, plot, and outcome, respectively.

Evaluation: We report accuracy (the official shared-task metric), precision, recall and F1 score on the development and test sets. For track A, we additionally compare against majority and random baselines. We do not have an additional evaluation for track B, as the embeddings used in track A are

[2] https://github.com/stanfordnlp/GloVe

[3] https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2

identical to track B and can only be evaluated on track A.

5 Results

Table 1 summarizes the results for both tracks. On the development set of track A, our system with SBERT embeddings reaches an accuracy of 0.6150 and F1 of 0.6244, outperforming the random baseline (Acc=0.5005) and the majority baseline in accuracy (Acc=0.5050). For the F1 score, the majority baseline is fairly strong and reaches a higher score than our system. Since the system with SBERT embeddings performed better on the development set, we used only this setup on the test set and to submit our official test labels to the shared task. On the test set for track A, the system achieves 60.25% accuracy and for track B, the official test result is 57.75% accuracy (see also the overall ranking in Hatzel et al. (2026)). The variant with GloVe embeddings does not perform higher than chance, giving it similar performance scores to the random baseline.

Overall, the results indicate that narrative similarity judgments can be supported by aspect-driven narrative representations: even a simple cosine-similarity decision rule based on our narrative embeddings improves over non-informed baselines on track A. At the same time, the performance leaves room for improvement, suggesting that both the elicitation of aspect text and the embedding construction may be critical levers for future work.

Table 2 summarizes the ablation experiments. The aspect pipeline slightly outperforms directly embedding the raw story summaries with the same MiniLM encoder (0.59 vs. 0.56 accuracy), suggesting that aspect elicitation adds useful structure beyond raw-text similarity. However, the gain is modest. The ablations show that themes are informative, but not uniformly beneficial: using only themes performs well, while dropping themes gives the best accuracy among the ablations. This supports our qualitative observation that broad thematic overlap can sometimes obscure event-level similarity. Simple manual weighting gives only small improvements, indicating that aspect combination remains a weakness of the current system.

When looking at the texts for the different aspects generated by the LLM, we see that it often is able to capture the important themes and plot points of a story in such a way that the sentence transformer embeddings were able to capture the similarities in the stories. For instance, for one triple, it generated

Triple 1

Anchor:

Themes:	"FAMILY", "SECRETS", "FRIENDSHIP", "REDEMPTION", "IDENTITY"
Plot:	"Sailor returns home after many years", "Helps friend revive struggling venue", "Discovers daughter is his own child", "Falls out with friend", "Saves friend from con man"
Outcome:	"Friendship is restored through redemption and truth."

Text A:

Themes:	"FAMILY", "SECRETS", "LOVE", "IDENTITY", "REGRET"
Plot:	"Independent woman reflects on her past", "First love returns after years", "Hides existence of their son", "Faces her concealed truth", "Revisits her fulfilled yet secretive life"
Outcome:	"Confrontation of secrets leads to personal growth."

Text B:

Themes:	"FAMILY", "SECRETS", "LOVE", "IDENTITY", "REDEMPTION"
Plot:	"Belgian wool dealer returns to homeland", "Single father leaves daughter behind", "Develops relationship with another single parent", "Family secrets are revealed", "Conclusion ties all stories together"
Outcome:	"Family bonds are strengthened through revealed truths."
Gold label:	Text B is closer to the anchor.
Prediction:	Text B is closer to the anchor.

However, sometimes overlap in the themes likely let the embeddings to be closer while the plot points and outcome suggest that the narratives are actually not similar. For example, for one triple, the LLM gave

Setting	Acc	F1	Precision	Recall
Devel A (GloVe)	0.4900	0.5000	0.4951	0.5050
Devel A (SBERT)	0.6150	0.6244	0.6154	0.6337
Devel A (majority)	0.5050	0.6711	0.5050	1.0000
Devel A (random)	0.5005	0.4971	0.5062	0.4901
Test A (SBERT)	0.6025	–	–	–
Test B (SBERT)	0.5775	–	–	–

Table 1: Results on the development and test set.

Setting	Acc	F1	Precision	Recall
Aspect pipeline	0.5897	0.5745	0.6067	0.5455
Raw-text MiniLM	0.5590	0.5743	0.5631	0.5859
Only themes	0.5949	0.5990	0.6020	0.5960
Only plot	0.5641	0.5854	0.5660	0.6061
Only outcome	0.5744	0.5654	0.5870	0.5455
Drop themes	0.6000	0.5938	0.6129	0.5758
Drop plot	0.5949	0.5949	0.6042	0.5859
Drop outcome	0.5949	0.6146	0.5943	0.6364
Best weighted	0.6000	0.5938	0.6129	0.5758

Table 2: Additional analyses on Devel A. "Best weighted" refers to simple manually tested Themes/Plot/Outcome weightings; the best accuracy was obtained by both $(0.2, 0.2, 0.6)$ and $(0.1, 0.2, 0.7)$.

Triple 2

Anchor:
Themes: "BULLYING", "POWER", "TRANSFORMATION", "ISOLATION", "HEROISM"
Plot: "Leo is bullied at new school", "Infects classmates with virus", "Learns to reverse the infection", "Becomes school's hero", "Gains power over peers"
Outcome: "Leo finds acceptance through manipulation."

Text A:
Themes: "ISOLATION", "IDENTITY", "MISUNDERSTANDING", "REVENGE", "CHANGE"
Plot: "Alberto is a work-focused employee", "Refuses contact with others", "Caught in misunderstandings", "People seek revenge on him", "Forces him to change identity"
Outcome: "Alberto is compelled to change."

Text B:
Themes: "ZOMBIES", "IMMUNITY", "VIOLENCE", "SURVIVAL", "TEAMWORK"
Plot: "Vaccine accidentally swapped with virus", "School turns into zombie chaos", "New student discovers swim team immunity", "Girls fight zombies with violence", "Team saves the day"
Outcome: "Swim team survives and restores order."

Gold label: Text B is closer to the anchor.
Prediction: Text A is closer to the anchor.

While the cosine similarity predicted that Text A is closer to the Anchor than Text B, the gold label says the opposite, likely due to both the Anchor and Text B being about infection outbreaks. While the LLM produces aspects that let one come to that conclusion, it produces themes for the Anchor and Text A that align more closely than the themes for the Anchor and Text B, likely leading the narrative embeddings to be more similar. This interpretation is supported by Table 2: themes are useful on their own, but dropping them slightly improves accuracy, suggesting that broad thematic overlap can sometimes override event-level similarity. We can also see here that the LLM ignored the request from the prompt to avoid highly story specific details like proper names, which also occurred frequently in other aspect generations.

The additional analyses also point to an imbalance in the aspect representations. Themes are short lists of isolated words, plot points are short event descriptions, and outcomes are single short phrases. Nevertheless, the sentence-transformer variant maps each aspect to the same 384-dimensional vector before concatenation. This equal dimensional treatment may not reflect equal informativeness. The ablation results suggest that the current representation does not fully exploit the three aspects and that learned weighting or attention over aspects may be more appropriate than simple concatenation.

6 Conclusion

We present a simple and interpretable pipeline for SemEval-2026 Task 4 that operationalizes narrativity via three aspects: themes, course of action, and outcomes. A prompted LLM produces constrained aspect texts, which are then converted into narrative embeddings using either static word embeddings with aspect averaging or aspect-wise sentence-transformer embeddings. The resulting narrative embeddings serve as the direct output for track B and enable cosine-similarity decisions for track A.

Our system improves over random and majority baselines in accuracy on the development set of track A and achieves 60.25% accuracy on the official test set. Additional analyses show that the aspect representation slightly outperforms directly embedding the raw story text, but that the gain is modest and that the three aspects contribute unevenly. However, it does not perform high in the shared task leader board, which can at least in part be attributed to issues with the consistency of the generated aspects and a high dependency on the quality of the embeddings encoding these aspects. Future work could improve robustness by refining the prompting strategy, for example through few-shot prompting, self-correction, or validation steps that remove proper names and enforce output constraints more reliably. The ablation and weighting results also suggest exploring stronger contemporary sentence encoders, learned aspect weights, attention over aspects, and task-specific modeling for Track A and Track B rather than using the same representation strategy for both.

References

H. Porter Abbott. 2014. Narrativity. In: the living handbook of narratology.

Hans Ole Hatzel, Ekaterina Artemova, Haimo Stiemer, Natalia Fedorova, Evelyn Gius, and Chris Biemann. 2025. Narrative similarity – annotation guidelines.

Hans Ole Hatzel, Ekaterina Artemova, Haimo Stiemer, Evelyn Gius, and Chris Biemann. 2026. SemEval-2026 Task 4: Narrative Similarity and Narrative Representation Learning. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, CA, USA.

Hans Ole Hatzel and Chris Biemann. 2023. Narrative cloze as a training objective: Towards modeling stories using narrative chain embeddings. In *Proceedings of the 5th Workshop on Narrative Understanding*, pages 118–127, Toronto, Canada. Association for Computational Linguistics.

Hans Ole Hatzel and Chris Biemann. 2024a. Story embeddings — narrative-focused representations of fictional stories. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 5931–5943, Miami, Florida, USA. Association for Computational Linguistics.

Hans Ole Hatzel and Chris Biemann. 2024b. Tell me again! a large-scale dataset of multiple summaries for the same story. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 15732–15741, Torino, Italia. ELRA and ICCL.

Hans Ole Hatzel, Fynn Petersen-Frey, Tim Fischer, and Chris Biemann. 2023. Dimensions of Similarity: Towards Interpretable Dimension-Based Text Similarity. In Kobi Gal, Ann Nowé, Grzegorz J. Nalepa, Roy Fairstein, and Roxana Rădulescu, editors, *Frontiers in Artificial Intelligence and Applications*. IOS Press.

Juri Opitz, Lucas Moeller, Andrianos Michail, Sebastian Padó, and Simon Clematide. 2025. Interpretable text embeddings and text similarity explanation: A survey. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 22303–22319, Suzhou, China. Association for Computational Linguistics.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.

Shashank Srivastava and Nebojsa Jojic. 2018. A spatial model for extracting and visualizing latent discourse structure in text. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2268–2277, Melbourne, Australia. Association for Computational Linguistics.

Igor Sterner, Alex Lascarides, and Frank Keller. 2026. Contrastive learning with narrative twins for modeling story salience. *Preprint*, arXiv:2601.07765.

Sean Wilner, Daniel Woolridge, and Madeleine Glick. 2021. Narrative embedding: Re-Contextualization through attention. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1393–1405, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

A Prompt template

```
You analyze short story/film plot summaries.

For EACH provided text, output:
1) themes: 1-5 themes, each ONE WORD,
    capitalized. Describes the defining
    constellation of problems, central ideas,
    and core motifs of a story. The definition
    does not cover the concrete setting of a
    story
2) plot: 1-5 plot points, each <= 10 words.
    Describes sequences of events, actions,
    conflicts, and turning points in a story and
     the order in which they happen
3) outcome: exactly 1 outcome, <= 10 words.
    Describes the results of the plot at the end
     of the text, for example, the conflict
    resolution, the characters' fates, moral
    lessons, etc. It does not cover intermediate
     statuses that change later in the story

Give similar descriptions for similar stories so
     that themes, plots and outcomes are
    comparable across summaries.
Avoid details that are specific to a story (e.g.
     avoid proper names from the story) and use
    generic themes, plot points and outcomes,
    ideally identically repeated across texts
    where it is fitting.

Important constraints:
- Themes: max 5 items, each exactly one word (no
     spaces/hyphens).
- Plot points: max 5 items, each <= 10 words.
- Outcome: string <= 10 words.
- Do NOT add any extra keys. Do NOT add
    commentary.
- Return STRICT JSON only, matching the provided
     schema.
```

Figure 2: Prompt used to elicit narrative aspects from an LLM.

B Original story summaries for the examples in Section 5

Triple 1:

```
{"anchor_text": "After many years away sailor
    Hannes Wedderkamp returns to his home city
    of Hamburg and meets up with his old friend
    Pittes Breuer who owns a venue on the
    Reeperbahn. Currently struggling, Hannes
    helps him turn it around by introducing a
    new revue. He also takes a kindly interest
    in his friend's grown-up daughter, only to
    discover that she is in fact really his own
    child who has been brought up by Pittes.
    When the two men fall out shortly afterwards
    , Pittes falls in with an unscrupulous con
    man and it falls to Hannes to save him.", "
    text_a": "Joan Verra is an independent,
    loving woman with a free and adventurous
    spirit. When her first love returns without
    warning after years of absence, she decides
    not to tell him that they had a son together
    . This lie by omission is an opportunity for
     her to revisit her life: her youth in
    Ireland, her professional success, her loves
     and her relationship with her son. A
    seemingly fulfilled life, but one which
    hides a secret that she will have to face.",
     "text_b": "Australia is about Edouard
    Pierson, a Belgian-born wool dealer who
    emigrated to Australia after World War II.
    The movie actual takes place in Belgium as
    he returns to his homeland to assist his
    family with their wool business. Edouard was
     left a single father after his girlfriend
    died and when he goes to Belgium he leaves
    behind this young girl, whom his family don'
    t know about. He meets a beautiful woman,
    Jeanne, another single parent, and an
    intense relationship develops. Edouard's
    relationship with his family has its ups and
     downs and many secrets are revealed before
    the movie's conclusion ties everything
    together."}
```

Triple 2:

```
{"anchor_text": "Having been a victim of
    bullying, life at his new school does not
    take a turn to the better for Leo Wei\u00df
    (played by Torge Oelrich (de). In order to
    no longer be the misfit and to exercise
    power over his fellow students, he secretly
    infects them with a zombie virus. As he is
    the only one who knows that a training in
```

dancing, handicrafts or mathematics can transform them back to normal, Leo soon becomes the school's highly acclaimed hero .", "text_a": "Alberto is an employee who is the Italian average of society of the Fifties. Alberto is a go-getter, attached only to his work, and believes that everyone meets him wants to bring Alberto bad luck. Alberto refuses every contact with other people, but soon finds himself caught in misunderstandings and so the people, to take revenge on him and his meanness, force him to change his identity.", "text_b": "A lab mix-up accidentally swaps a vaccine with a virus that turns a high school full of students and teachers into flesh-eating zombies. But all is not lost: New student Aki discovers that the swim team is immune to the plague. With the school rampaged by ravenous monsters, the girls engage in an over-the-top orgy of gory violence to save the day. Sasa Handa, Yuria Hidaka and Hiromitsu Kiba star in this comic creature feature."}

Triple 3:

{"anchor_text": "This cinematic adaptation of the autobiography of Anna Wimschneider depicts her life's experiences and workaday routines as a woman born on a farm in Lower Bavaria, Germany in the 1920s. Anna's mother died young in childbirth and Anna had to take her place and work very hard. On a Nazi Party rally she meets young Albert, who owns a farm. They realize that they both don't believe in fascism and go to a coffee bar where he starts wooing her. Against her prior decision to leave farm life as soon as possible, she agrees to marry him, hoping that her life will become easier on Albert's farm.", "text_a": "Half Broke Horses is the story of Lily Casey Smith's life. Author Jeannette Walls, the granddaughter of Lily Casey Smith, wrote the book from Lily's perspective.\nAs a child growing up on the frontier in Texas, Lily learns how to break horses. At the age of fifteen, she rode five hundred miles, alone, to get to her job as a teacher in a one-room schoolhouse. Later in her life, Lily runs a vast cattle ranch in Arizona, along with her second husband and their two children. A woman of many talents, Lily earns extra money at various points in her life by playing poker, selling bootleg liquor, and riding in horse races.\nHalf Broke Horses depicts the freedom of rural life, its joys and struggles, and celebrates the courage and spirit of its protagonist, Lily Casey Smith. Walls says the book is \u201cin the vein of an oral history, a retelling of stories handed down by my family through the years, and undertaken with the storyteller\u2019s traditional liberties.\u201d", "text_b": "Jane Falbury (Judy Garland) is a farm owner whose actress sister, Abigail (Gloria DeHaven), arrives at the family farm with her theater troupe. They need a place to rehearse, and Jane and her housekeeper, Esme (Marjorie Main), reluctantly agree to let them use their barn. The actors and actresses, including the director, Joe Ross (Gene Kelly), repay her hospitality by doing chores around the farm. Although Joe is engaged to Abigail, he begins to fall in love with Jane after Abigail leaves him in an angry fit. Similarly, although Jane is engaged to Orville (Eddie Bracken), she falls in love with Joe."}