Annotation Guideline

1. Task Definition

Traditional sentiment analysis focuses on the aspect of a target (e.g., battery performance in electronic products). Opinion analysis typically extracts Holder, Target, and Expression. Events, however, convey richer information compared to nominal aspects and are prevalent on social media. Prior event-based sentiment analysis often classifies sentiments too coarsely.

To refine granularity, we propose a joint task of event-opinion analysis, extracting the following pairs and triples:

Pair

```
<Trigger, Argument 1, Argument 2, …, Argument n>
```

Triplet

```
<Trigger, Opinion Polarity, Opinion Expression>
```

These structures represent the sentiment polarity and expression regarding events. The relationship between event-opinion analysis and cross-document scenarios includes:

1. Different users may express conflicting opinions on the same event across posts, requiring cross-document extraction for event information completion.

2. In some cases, posts from the same user on an event may lack explicit opinion expressions, which cross-document extraction can address.

2. Annotation Scheme

Elements and Types

Element	Type	Example
Event Argument	Person	Trump
	Organization	Tesla
	Country	China, USA
	Location	China, New York
	Date	16th, September 2023
	Object	Products, tools
	Other	Other types not listed
Trigger Word	Trigger	Prohibit, Lead
Opinion Polarity	Sentiment##Support	Support, Oppose, Neutral
	Sentiment##Negative	
	Sentiment##Neutral	
Opinion Expression	Clause with sentiment	For sentences containing opinion polarity, mark the entire sentence (not just a single word)
Additional Annotation	Field##Sports	
	Field##Society	
	Field##Technology	
	Field##Entertainment	
	Field##Politics	
	Field##Finance	
	Field##Disaster	
	Field##Business	
	Field##Military	
	Field##Else	

Key Annotation Rules

1. **Event Arguments**: A single word, including seven types: Person, Organization, Country, Location, Date, Object, and Other.

2. **Triggers**: A single word, without a specific category, usually a verb. If present, mark it directly.

3. **Connection Between Triggers and Event Arguments**: Due to the diversity of hot news events on social media, predefining event categories and specific relationships between triggers and arguments is highly time-consuming. Therefore, we plan to construct an open event-opinion joint extraction dataset, focusing on extracting as many event structures as possible from the text rather than adhering to predefined fixed relationships. **Triggers and their related event arguments are uniformly connected using a unique relationship identifier, "Associate,"** to facilitate the construction of event-opinion quadruples later.

4. **Opinion Polarity**: Classified as Support, Oppose, or Neutral based on sentiment.

5. **Opinion Expression**: Entire sentiment-expressing sentence is annotated.

Relationships:

- `Associate` : Event triggers and their corresponding event arguments; sub-events and their corresponding opinions when sub-events exist.

- `Sub-event` : The trigger word of the main event and the trigger word of the sub-event.

3. Data Filtering

Since the data undergoes manual pre-screening, some texts may be deemed unsuitable for annotation. After discussion, the **texts to be excluded** are categorized as follows:

1. Issues with information substitution (e.g., "chen Yunshan").

2. Discussions about event topics, typically in the form of questions.

3. Government work reports or personal statements (e.g., the 14th Party Congress speeches are retained but not annotated for now).

4. Discussions on topics related to the event's subject following a hot event. These may include emotional tendencies or opinions but do not supplement the event's completeness. Such topics often derive from the event, like debates on morality or legal responsibility.

Special Cases Requiring Annotation:

1. **Comments supplementing event causes or specific details but lacking emotional tendencies or opinions**:

   - **Neutral Polarity Annotation**:

     - Case 1: Sentences without opinion expression should not be annotated.

     - Case 2: Sentences with opinion expression but neutral in sentiment should be annotated with a neutral label.

2. **Events containing multiple related sub-events**:

   - All relevant sub-events should be annotated.

4. Examples

1. Politics

环球网报道 综合美国有线电视新闻网（CNN）、《华盛顿邮报》等多家媒体报道，负责主审美国前总统特朗普涉嫌财务欺诈案的纽约法官亚瑟恩戈隆当地时间16日作出裁定，内容包括要求特朗普以及特朗普集团支付超过3.5亿美元的罚款、禁止特朗普在3年内担任纽约任何公司的董事。对于该裁定，特朗普当天晚些时候称将上诉。据CNN报道，特朗普16日在海湖庄园发表讲话时称，这是"非常悲伤的一天"。"我们会上诉，我认为我们会成功。"他补充称。据媒体此前报道，2023年9月26日，纽约曼哈顿一家法院的法官亚瑟恩戈隆裁定纽约州总检察长利蒂希娅詹姆斯对特朗普等人及特朗普集团为获取经济利益而夸大其房地产资产估值等财务欺诈行为的核心指控有效。法院自同年10月2日起开始审理詹姆斯对特朗普提出的其他六项指控。詹姆斯寻求对特朗普等人及特朗普集团处以罚款并禁止他们在纽约州经营企业等。<sep>当地时间2月16日，据彭博社消息，美国前总统特朗普因夸大其净资产构成欺诈，被处以超3.5亿美元罚款。美国纽约州法官当日对美国前总统特朗普的民事欺诈案作出裁决。由于特朗普将其净资产夸大了数十亿美元，构成了严重欺诈行为，法官裁定对特朗普处以超3.5亿美元的罚款，永久禁止特朗普进入纽约房地产行业，并大幅限制他在该州开展业务的能力。此外，法官还裁定对特朗普的两个儿子小唐纳德特朗普和埃里克特朗普实施五年的行业禁令。（央视新闻客户端）<sep>看起来是吹牛也要上税，实际上是党争，是政治迫害。<sep>他也不只是吹牛被罚吧。吹牛的时候欺骗银行，抵押贷款。报税时，又降低财产，偷税。<sep>哇哦，美国干的不错哦<sep>在美国吹牛要罚款啊？我们的地产商如果不能吹牛，这个行业估计没人了。<sep>吹牛有风险！，理由是财务欺诈，夸大自家企业净资产，以获取不当利益，翻译过来就是做广子吹牛吹破了。，要上诉，能赢。<sep>反正不用花特朗普的钱，特朗普支持者们赶紧捐款吧。<sep>像他会干出来的事 脑子一直都不太好 被罚也是应得的<sep>在美国吹牛皮是犯法的，还要上税的；懂王还被永久被禁止进入纽约房地产行业<sep>并不影响啊，执行可以拖，拖到他当选总统以后，赦免自己就行了<sep>传下去！美国那边吹牛逼犯法！！！吹牛逼要上税！前总统因为吹牛逼被罚了3.5个亿美元！<sep>判了，构成严重欺诈行为！特朗普被罚超3.5亿美元，永久禁止进入纽约房地产行业，并对其两个儿子实施五年的行业禁令。(新华每日电讯)<sep>下图判决是个好消息！这货1月公开讲，他当选了要对某国所有商品实施60重关税。转：刚

Pairs and Triples:

Example 1:

Pair:

```
<Argument: Arthur Engoron, 16th | Trigger: Ruling>
<Argument: Trump, Trump Organization, Fine | Trigger: Pay>
<Argument: Arthur Engoron | Trigger: Prohibit>
<Argument: Trump, Director of any New York company | Trigger:
Hold>
<Argument: Arthur Engoron, Donald Trump Jr., Eric Trump, Five
Years, Industry Ban | Trigger: Implement>
```

Triple:

```
<Trigger: Ruling | Polarity: Oppose | Opinion: "This is not a
bout taxes, but party strife and political persecution.">
```

Example 2:

Pair:

```
<Argument: Arthur Engoron, 16th | Trigger: Ruling>
<Argument: Trump, Trump Organization, Fine | Trigger: Pay>
<Argument: Arthur Engoron | Trigger: Prohibit>
<Argument: Trump, Director of any New York company | Trigger:
Hold>
<Argument: Arthur Engoron, Donald Trump Jr., Eric Trump, Five
Years, Industry Ban | Trigger: Implement>
```

Triple:

```
<Trigger: Ruling | Polarity: Neutral | Opinion: "It's not jus
t about boasting and being fined, is it?">
```

Example 3:

Pair:

```
<Argument: Arthur Engoron, 16th | Trigger: Ruling>
<Argument: Trump, Trump Organization, Fine | Trigger: Pay>
<Argument: Arthur Engoron | Trigger: Prohibit>
<Argument: Trump, Director of any New York company | Trigger:
Hold>
<Argument: Arthur Engoron, Donald Trump Jr., Eric Trump, Five
Years, Industry Ban | Trigger: Implement>
```

Triple:

```
<Trigger: Ruling | Polarity: Support | Opinion: "Wow, the U.
S. did a great job.">
```

Example 4:

Pair:

```
<Argument: Arthur Engoron, 16th | Trigger: Ruling>
<Argument: Trump, Trump Organization, Fine | Trigger: Pay>
```

```
<Argument: Arthur Engoron | Trigger: Prohibit>
<Argument: Trump, Director of any New York company | Trigger:
Hold>
<Argument: Arthur Engoron, Donald Trump Jr., Eric Trump, Five
Years, Industry Ban | Trigger: Implement>
```

Triple:

```
<Trigger: Ruling | Polarity: Support | Opinion: "Being fined
is well-deserved.">
```

......

5. Issues

1. **Nested Events**

   Many events include multiple sub-events. For example, in the first case:

   ```
   <Argument: Arthur Engoron, 16th | Trigger: Ruling>
   <Argument: Trump, Trump Organization, Fine | Trigger: Pay>
   ```

   This sentence contains two nested events, where the second event ("Pay") is the target content of the trigger word "Ruling" in the first event.

   Solution:

   For two (or more) nested events, annotate the relationship between the triggers using the label "sub-event."

2. **Defining Opinion Boundaries**

   When annotating user opinions, should we identify the smallest sentence containing the opinion polarity or the largest? For instance:

   Event:

```
<Argument: SSE 50, CSI 300 dividend yield | Trigger: Appro
aching historical high>
```

User opinion:

> "For the sake of sensationalism, they really have no shame. Everyone is at the bottom of the class but still feels proud of themselves."
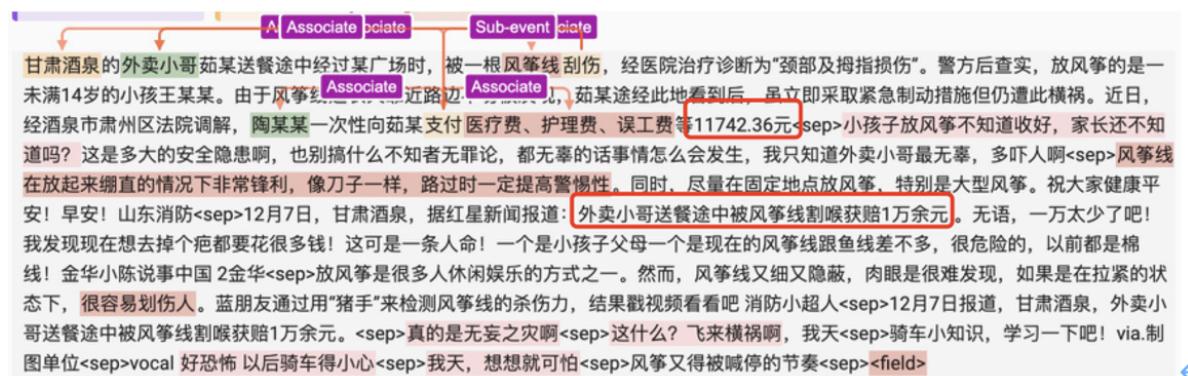
Should the entire statement be annotated, or only a portion (e.g., the first two sentences, which have stronger polarity)?

Solution:

The boundaries of opinions are difficult to define. Annotation can be based on individual judgment. During model evaluation, adopt two strategies:

- **Restrict**: The opinion boundary must exactly match the gold standard.
- **Loose**: If the identified opinion overlaps with the gold standard, it is considered correct.

3. **Do supplementary sub-events mentioned in comments, which have already been annotated as existing sub-events, need to be annotated again?**

For instance, in the event described above, the comment "The deliveryman was compensated over 10,000 yuan after his throat was cut by a kite string while delivering food" has already been detailed in the original news article

(the first text). Should this event description in the comments be annotated if it does not provide missing information about the original event but rather summarizes or restates it?

Solution: If the supplementary sub-event in the comment is merely a repetition or summary of an already annotated sub-event from the original text and does not provide additional information, it does not need to be annotated again. However, if the comment adds new details or context to the sub-event, it should be annotated.

4. **Sentiment Polarity for Comments on Accidental Events**

   Comments on accidental events often express regret. How should the sentiment polarity for such events be annotated?

   Solution:

   Annotate the sentiment polarity as **negative**.

5. **Sentiments and Events: Handling Sub-Events**

   The opinion is not directed at the main event of "publishing a reward notice" but instead at someone in the notice going viral due to their appearance. How should emotions and events be selected?

红星新闻据内蒙古广播电视台、九派新闻、大皖新闻等多家媒体报道，近日，呼和浩特市回民区人民法院发布悬赏公告，寻找被执行人曹晓娜。曹晓娜出生于2004年，涉案标的为10028元。悬赏公告中，年仅19岁的被执行人曹晓娜照片容貌清秀，被网民称为"高颜值女老赖"后引发关注。12月9日中午，"回民区法院发布"微信公众号发布《呼和浩特市回民区人民法院关于申请执行人杨某喜与被执行人曹某娜房屋租赁合同纠纷一案结案公告》，《公告》显示，关于申请执行人杨某喜与被执行人曹某娜（2023）内0103执恢994号房屋租赁合同纠纷一案，被执行人曹某娜于2023年12月8日履行全部涉案款项，该案执行完毕，现已结案。"回民区法院发布"微信公众号、"呼和浩特市回民区人民法院"抖音账号时常发布当地法院的悬赏公告。12月9日中午，红星新闻记者检索这两个经常发布悬赏公告的账号发现，12月7日当天有多条悬赏公告发布，但已经无法查看到曹晓娜的悬赏公告。12月9日，红星新闻记者通过中国执行信息公开网查询发现，曹晓娜在房屋租赁合同纠纷一案中未按执行通知书指定的期间履行生效法律文书确定的给付义务，执行标的为10028元。中国执行信息公开网信息显示，该房屋租赁合同纠纷案于2023年10月8日立案，终本日期为2023年11月27日，即11月27日是终结本次执行程序的日期。限制消费令显示，曹晓娜被限制消费的内容包括旅游、度假等。但在网曝的疑似曹晓娜本人账号中，12月6日该账号仍在分享呼和浩特大青山的滑雪照片。红星新闻记者检索发现，该网曝疑似曹晓娜本人账号与曹晓娜本人同名，且有曾在今年11月8日发布过的悬赏公告中所使用的曹晓娜照片。12月9日，红星新闻记者致电悬赏通告上的承办人联系电话，截至发稿时并未拨通。<sep>19岁女子上悬赏公告，因"高颜值"受关注。总共就1万多块钱，开个直播就赚回来了，至于吗？是不是炒作啊？<sep>油饼？老赖就是老赖，还管长的好不好看？一个破老赖还吹上高颜值了?什么巨心?谁在乎她长什么样子？？想想疯了???想说明什么？三观跟着五官走吗<sep>这个热搜体现了什么样的价值观？<sep>有些人真的跪久了老赖就是老赖还扯什么高颜值老赖。老色批就是老色批没有帅哥老色批。<sep>啊？我没看错吧？这是被执行人诶 三观跟着五官走 也不能这么走吧 况且这五官也没有那么够吧 吃饱了撑的还是看不到美女了不理解 理解不了一点 虽然已经还清了 但是也<sep>虽然但是嗯 这个时候也不能光看颜值呀<sep>已经悬赏了 谁会去在乎她的颜值<sep>三餐茶饭，四季衣裳，共同造就了一个叫家的地方<sep>曹晓娜，生于2004年8月12日，肖猴，名有不周也，一来名字欠缺补益性，二来名字没有安全性，因为肖猴者名字慎娜忌晓：<sep>一万块上热搜榜。深圳这边有五六百人被健身房诈骗150多万警方还不给立案调查的事，哪个媒体也不愿意报道。<sep>这个世界已经这么疯了吗 长得好看的还不够你喜欢吗？果然三观跟着五官走<sep><field>

Personal View:

It has already been decided in previous discussions to annotate sub-events. As long as an event is relevant to the main event, it should be annotated. For example, "publishing a reward notice" as the main event is incomplete without sub-events, such as describing the content of the notice (e.g., sentencing or penalties). If sub-events are not annotated, the information about the entire event would be incomplete. Therefore, **annotating sub-events is necessary.**

Answer:

Yes, it can be annotated.

If a user expresses an opinion about a sub-event, such as in the following example:

- "Restarting the livestream" is the main event, while "the livestream room was banned" is a sub-event. Users are expressing opinions about the sub-event "being banned" rather than the main event "restarting the livestream."

If opinions are annotated, the sentiment is directed at the sub-event. In some cases, user opinions on the main event and sub-event may differ. For example:

- Sub-event: "The livestream room was banned" (sentiment: **support**).

- Main event: "Restarting the livestream" (no explicit sentiment polarity, but it can be inferred that users are opposed to it, sentiment: **positive**).

Solution:

A link relationship can be added between opinions and events to associate them. For sentiment polarity, only consider explicit opinions for now. Implicit opinions requiring inference are difficult to annotate and challenging for machines to recognize.

7. Determining Sentiment for Criticism of Partial Arguments in an Event

Example:

The main event is "Ms. Huang reports her ex-husband," while user comments mostly criticize the behavior of the ex-husband.

Personal View:

If users express opinions on the main event (or both the main event and sub-events), annotate the sentiment polarity directly based on their expression regarding the main event. If the opinion targets a sub-event, further discussion is required to decide how to annotate it.

Annotation Approach:

The annotation approach is the same as previously described.

8. Determining Sentiment Polarity When Comments Target a Sub-Event

Question:

When an event contains multiple sub-events, and user comments focus on a specific sub-event, how should the sentiment polarity of the opinion be determined? Is it possible to infer the sentiment polarity for the entire event based on the sub-event's polarity?

，检方以故意伤害罪提起公诉 12月6日，辽宁省沈阳市XXXXXXXXX起诉书。1965年出生的男子高某某因怀疑54岁的杨某某在与其XXXXXXXXXX正XXXXX系，曾多次对其进行殴打，用燃烧的烟蒂将其烫伤。2023年1月17日至18日，高某某持XXXXXXXX布鞋殴打杨某某，致其于2023年1月19日死亡。经鉴定，杨某某系因多次钝性外力作用，造成周身广泛软组织损伤，导致创伤性休克而死亡。2023年1月20日，高某某经公安机关传唤到案。检方认为，高某某故意伤害他人身体，致一人死亡，应当以故意伤害罪追究其刑事责任。（九派新闻）<sep>请问有什么大吗? 这种人就应该直接抓进去坐牢，别到时候用什么精神病来做挡箭牌！<sep>标题能不能把犯罪人写在前面? 为什么要强调死者的年龄而不强调犯罪事实? <sep>是不是得单身才能保平安啊，真的好可怕啊，现在动不动就置人于死地<sep>每年那么多家暴案，纱妻案，但你一说因为这些害怕男的害怕婚姻就被骂极端<sep>男方早就开始对女方进行长期家暴了吧，标题却在这强调女方出轨，请问是什么意思呢，家暴男就不应该活在这个世界上<sep>故意伤害罪<sep>男的神金冰啊还把人家打三了<sep>就是家暴男找借口，标题觉呵呵新闻学<sep>不懂为啥把逝者放前面，这个男的打人致死耶，避重就轻?而且一切只是怀疑，不能好好沟通?占有欲这么强?<sep>一把年纪还这么不理智 这下好了 往后余生就在里面过吧还有这标题能不能尊重一下逝者呢<sep>这也太吓人了！！女孩们都擦亮眼睛吧<sep><field>

Example:

- **Main Event**: "The court initiated prosecution."

- **Sub-Events**:

  1. "Gao assaulted Yang."

  2. "Yang died from traumatic shock."
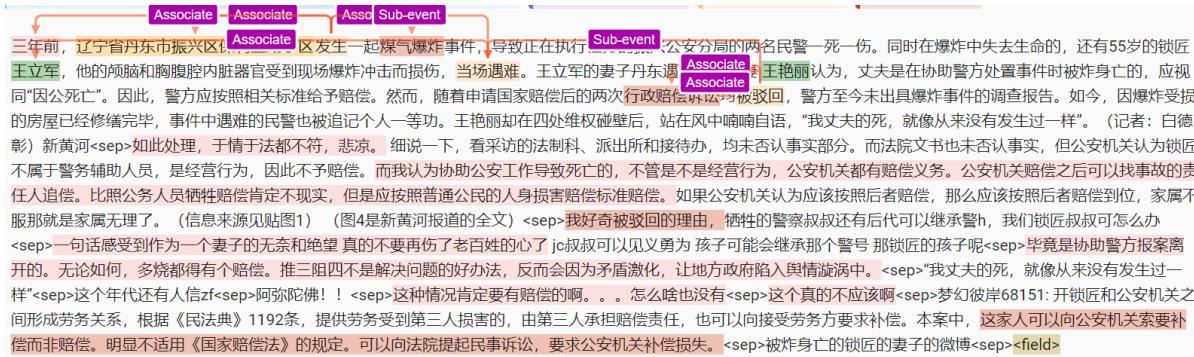
User comments typically evaluate the actions of the individual (e.g., "Gao") rather than the main event (e.g., the court's prosecution). For instance:

- A comment like "They should just throw him in jail" could indicate support for the court's decision to prosecute.

Annotation Approach:

The annotation approach is the same as previously discussed:

- Use the sentiment polarity of the sub-event to infer the sentiment towards the main event.

- Link opinions to the trigger words of the corresponding events to establish the relationship between them.

- Annotate the inferred sentiment for the main event (e.g., **support** in this example).

三年前，辽宁省丹东市振兴区○○区发生一起煤气爆炸事件，导致正在执行○○○公安分局的两名民警一死一伤。同时在爆炸中失去生命的，还有55岁的锁匠王立军，他的颅脑和胸腹腔内脏器官受到现场爆炸冲击而损伤，当场遇难。王立军的妻子丹东遇○○王艳丽认为，丈夫是在协助警方处置事件时被炸身亡的，应视同"因公死亡"。因此，警方应按照相关标准给予赔偿。然而，随着申请国家赔偿后的两次行政赔偿诉讼均被驳回，警方至今未出具爆炸事件的调查报告。如今，因爆炸受损的房屋已经修缮完毕，事件中遇难的民警也被追记个人一等功。王艳丽却在四处维权碰壁后，站在风中喃喃自语，"我丈夫的死，就像从来没有发生过一样"。（记者：白德彰）新黄河<sep>如此处理，于情于法都不符，悲凉。细说一下，看采访的法制科、派出所和接待办，均未否认事实部分。而法院文书也未否认事实，但公安机关认为锁匠不属于警务辅助人员，是经营行为，因此予赔偿。而我认为协助公安工作导致死亡的，不管是不是经营行为，公安机关都有赔偿义务。公安机关赔偿之后可以找事故的责任人追偿。比照公务人员牺牲赔偿肯定不现实，但是按照普通公民的人身损害赔偿标准赔偿。如果公安机关认为应该按照后者赔偿，那么应该按照后者赔偿到位，家属不服那就是家属无理了。（信息来源见图1）（图4是新黄河报道的全文）<sep>我好奇被驳回的理由，牺牲的警察叔叔还有后代可以继承警h，我们锁匠叔叔可怎么办<sep>一句话感受到作为一个妻子的无奈和绝望 真的不要再伤了老百姓的心了 jc叔叔可以见义勇为 孩子可能会继承那个警号 那锁匠的孩子呢<sep>毕竟是协助警方报案离开的。无论如何，多烧都得有个赔偿。推三阻四不是解决问题的好办法，反而会因为矛盾激化，让地方政府陷入舆情漩涡中。<sep>"我丈夫的死，就像从来没有发生过一样"<sep>这个年代还有人信zf<sep>阿弥陀佛！！<sep>这种情况肯定要有赔偿的啊。。。怎么啥也没有<sep>这个真的不应该啊<sep>梦幻彼岸68151: 开锁匠和公安机关之间形成劳务关系，根据《民法典》1192条，提供劳务受到第三人损害的，由第三人承担赔偿责任，也可以向接受劳务方要求补偿。本案中，这家人可以向公安机关索要补偿而非赔偿。明显不适用《国家赔偿法》的规定。可以向法院提起民事诉讼，要求公安机关补偿损失。<sep>被炸身亡的锁匠的妻子的微博<sep><field>

Additional Example:

- **Main Event**: "A gas explosion occurred."

- **Sub-Events**:

    1. "Wang Lijun died on the spot."

    2. "The administrative compensation lawsuit was dismissed."

User comments are directed at the sub-event "The administrative compensation lawsuit was dismissed," while the main event is an objective disaster. In this case, the primary focus of the entire event chain is the sub-event "The administrative compensation lawsuit was dismissed."

Annotation:

The sentiment polarity should be determined based on the user's opinion regarding the sub-event "The administrative compensation lawsuit was dismissed."

9. **Temporal Arguments**:
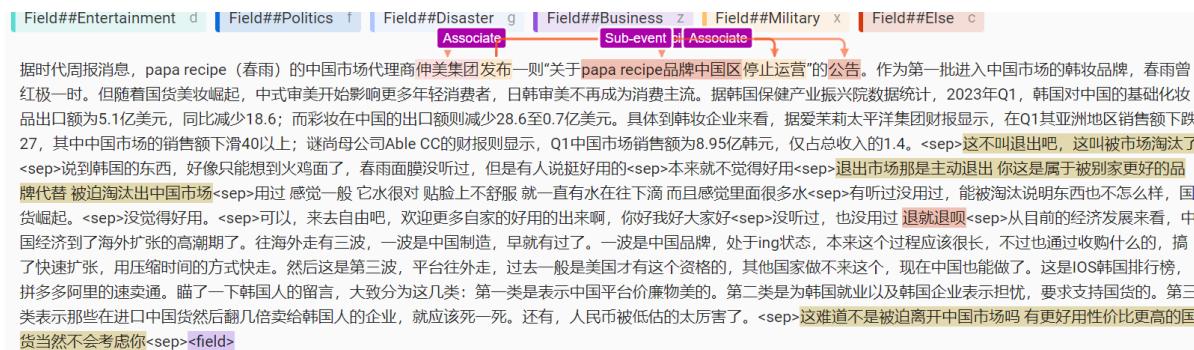
For ambiguous time expressions like "recently," should they be annotated?

Answer:

- Exact amounts (e.g., "11742.36 yuan," "over ten thousand yuan") can be annotated.

- Events should be annotated with precise temporal information, while ambiguous time expressions should not be annotated.

10. **Difficult-to-Determine Sentiment Polarity**:

For phrases like "Shenmei Group issued a statement" and "exited the Chinese market," user opinions often sarcastically imply that the announcement was made out of necessity. Should the sentiment polarity be annotated as "support" or "oppose"?

Answer:

- Sentiment polarity should be determined based on the result. Since users generally agree with the company's exit from the Chinese market, the sentiment should be annotated as **"support"** even if the semantics feel counterintuitive.

11. Scope of trigger words: In "An incident of a gas explosion occurred", "occurred" is the trigger word, while "gas explosion" can be considered as the event argument. If "an incident of" is removed, resulting in "A gas explosion occurred", does the trigger word change to " gas explosion occurred" instead of just "occurred"?

    Without considering discontinuous trigger words and given that "occurred" is a light verb, it is not used alone as an event trigger. In this example, it is possible to label either " gas explosion occurred" or "gas explosion".

12. "During the period from January 1, 2019, to October 21, 2021," treat "from January 1, 2019, to October 21, 2021" as a single event argument (temporal entity) for annotation.