

## Subtypes of Biographical Relevance

In the following, we provide a definition for the subtypes of biographical relevance that we provide as a second annotation layer in our dataset. Each of these subtypes can only be assigned to a tweet that has been labeled as conveying biographical relevance in the first annotation layer. The subtypes are **not** mutually exclusive. This means that it may be possible that several of the following subtypes apply for the same tweet. In this case, the annotator should assign all these subtypes simultaneously to the tweet.

- **General Likes**

We use this subtype for any of the author's non-trivial likes. Trivial likes are likes that hardly characterize the respective individual human. For example, if the author expresses his fondness about nice weather this hardly qualifies as a non-trivial like since virtually all humans like nice weather. (Actually, such trivial likes will not have been considered as biographically relevant in the first place.)

- **General Dislikes**

We use this subtype for any of the author's non-trivial dislikes. Trivial dislikes are dislikes that hardly characterize the respective individual human. For example, if the author expresses his dislike against unfriendly people this hardly qualifies as a non-trivial dislike since virtually all humans dislike unfriendly people. (Actually, such trivial dislikes will not have been considered as biographically relevant in the first place.)

- **Desires and Needs**

This subtype is used for any tweet in which the author expresses some desire or need. Further, the respective desire or need should characterize the author of the tweet. For example, if the author states that they are hungry then this is only a temporary need which may also apply to any living being. (Actually, such non-trivial/temporary desires and needs will not have been considered as biographically relevant in the first place.)

- **Food and Drink Preferences**

This subtype is used for tweets in which the author expresses their food and drink preferences. However, not every tweet in which the author talks about food/drink consumption qualifies for this subtype. It only qualifies for this subtype if the utterance tells us something about the author. That is, it should be clear that the author consumes the type of food on a regular basis or the type of food is special (for example, eating caviar is special and it suggests that the author is rich). If the author reports on eating a sandwich then this hardly qualifies for this category.

- **Job and Education**

By this subtype, we subsume all types of information about the author's education (e.g. which type of school/university they have attended) and professional background.

- **Family Relations**

By this subtype, we subsume family relationships and all types of friendships (both present and past) that involve the author of the tweet. We also use this label if a tweet reveals that the author of the tweet is/was the owner of some pet.

- **Habits**

By this subtype, we include any habit of the author as long as it is non-trivial. For example, buying food in the supermarket on a regular basis does not qualify since this is some habit that should apply to (almost) every adult human. (Actually, such trivial habits will not have been considered as biographically relevant in the first place.)

- **Personality Traits**

The subtype is assigned to tweets in which the author reflects on his personality. This subtype is not used for emotional states as long as they are only temporary. For example, a temporary sadness will not qualify but a tendency towards melancholy will. (Actually, temporary emotional states will not have been considered as biographically relevant in the first place.)

- **Ideological Views**

This subtype is used if the author expresses some of their ideological views, i.e. views on society, politics, economy or religion. The membership of some political party may also qualify for this subtype.

- **Health**

By this subtype we subsume all tweets in which the author reveals some information about their health. Notice that common/temporary health issues, such as having a cold, do not qualify for this subtype. (Common/temporary health issues will actually not have been considered as biographically relevant in the first place.)

- **Life Events**

We are looking for major life events which are defined as events that are environmental circumstances that have an identifiable onset and ending and may carry the potential for altering an individual's present state of mental or physical well-being. This subtype subsumes all major life events, such as graduation, marriage, divorce, becoming a parent, loss of job, being promoted, retiring etc.

- **Confidential Information**

This subtype includes all the biographical information about the author of a tweet that they should not disclose publicly on Twitter (for various reasons), such as liquidity status, passwords, phone number, exact place of residence, private email address, bank account number etc. This subtype is also used for any information that could in some way be used to the author's disadvantage. This includes information about consumption of illegal drugs or the author's confession of some addiction.

- **Other Demographic Information**

This subtype should be used for any (non-confidential) demographic information about the author that does not fit to the previous categories. Typical relations that qualify for this category are: height, weight, age, place of birth, eye colour, hair colour, nationality etc.<sup>1</sup>

- **Other**

While we tried to come up with an inventory of subtypes that should characterize the different forms of biographically relevance in our tweets in the best possible way, this inventory cannot be exhaustive. All instances of biographically relevance which cannot be assigned to any of the above categories will be assigned the subtype *Other*.

---

<sup>1</sup> We deliberately had some separate individual category for the demographic information *Family Relations* and *Job and Education* because these represent the subset of demographic information which we observed most frequently on our data and therefore decided that this should be acknowledged with a dedicated subtype of biographical relevance.