

Analyzing The 140 Characters

The online documentation is always the definitive source for Twitter platform objects, and it's worthwhile to bookmark the Tweets page, because it's one that you'll refer to quite frequently as you get familiarized with the basic anatomy of a tweet. No attempt is made here or elsewhere in the book to regurgitate online documentation, but a few notes are of interest given that you might still be a bit overwhelmed by the 5 KB of information that a tweet comprises. For simplicity of nomenclature, let's assume that we've extracted a single tweet from the search results and stored it in a variable named `t`. For example, `t.keys()` returns the top-level fields for the tweet and `t['id']` accesses the identifier of the tweet.

- The human-readable text of a tweet is available through `t['text']`:
RT @hassanmusician: #MentionSomeoneImportantForYou God.
- The entities in the text of a tweet are conveniently processed for you and available through `t['entities']`:

```
{
  "user_mentions": [
    {
      "indices": [
        3,
        18
      ],
      "screen_name": "hassanmusician",
      "id": 56259379,
      "name": "Download the NEW LP!",
      "id_str": "56259379"
    }
  ],
  "hashtags": [
    {
      "indices": [
        20,
        50
      ],
      "text": "MentionSomeoneImportantForYou"
    }
  ],
  "urls": []
}
```

- Clues as to the “interestingness” of a tweet are available through `t['favorite_count']` and `t['retweet_count']`, which return the number of times it's been bookmarked or retweeted, respectively.
- If a tweet has been retweeted, the `t['retweeted_status']` field provides significant detail about the original tweet itself and its author. Keep in mind that sometimes the text of a tweet changes as it is retweeted, as users add reactions or otherwise manipulate the text.

- The `t['retweeted']` field denotes whether or not the authenticated user (via an authorized application) has retweeted this particular tweet. Fields that vary depending upon the point of view of the particular user are denoted in Twitter's developer documentation as *perspectival*, which means that their values will vary depending upon the perspective of the user.
- Additionally, note that only original tweets are retweeted from the standpoint of the API and information management. Thus, the `retweet_count` reflects the total number of times that the original tweet has been retweeted and should reflect the same value in both the original tweet and all subsequent retweets. In other words, retweets aren't retweeted. It may be a bit counterintuitive at first, but if you think you're retweeting a retweet, you're actually just retweeting the original tweet that you were exposed to through a proxy. See *Examining Patterns in Retweets* later in this chapter for a more nuanced discussion about the difference between retweeting vs quoting a tweet.

You should tinker around with the sample tweet and consult the documentation to clarify any lingering questions you might have before moving forward. A good working knowledge of a tweet's anatomy is critical to effectively mining Twitter data.