# Rohdaten

Die Dateien sind pro Forum enthalten, wobei der Forumsnamen gleichzeitig Verzeichnisnamen ist, und die Klasse im Single-Label-Classifier-Fall darstellt.

## Posts.xml

Liste aller Posts. Dies sind die Elemente, die es eigentlich gilt zu Klassifizieren.

<row Id="1" PostTypeId="1" AcceptedAnswerId="925" CreationDate="2010-07-09T19:07:45.460" Score="59" ViewCount="23590" Body="**&lt;p&gt;My chocolate chips cookies are always too crisp. How can I get chewy cookies, like those of Starbucks?&lt;/p&gt;&#xA;&#xA;&lt;hr&gt;&#xA;&#xA;&lt;p&gt;Thank you to everyone who has answered. So far the tip that had the biggest impact was to chill and rest the dough, however I also increased the brown sugar ratio and increased a bit the butter. Also adding maple syrup helped. &lt;/p&gt;&#xA;"** OwnerUserId="16" LastEditorUserId="41" LastEditDate="2011-08-22T23:44:41.317" LastActivityDate="2016-12-26T07:46:32.293" Title="**How can I get chewy chocolate chip cookies?**" Tags**="&lt;baking&gt;&lt;cookies&gt;&lt;texture&gt;**" AnswerCount="10" CommentCount="5" FavoriteCount="22" />

<row Id="2" PostTypeId="1" AcceptedAnswerId="26" CreationDate="2010-07-09T19:08:34.570" Score="29" ViewCount="113148" Body**="&lt;p&gt;I've heard of people cooking bacon in an oven by laying the strips out on a cookie sheet. When using this method, how long should I cook the bacon for, and at what temperature?&lt;/p&gt;&#xA;**" OwnerUserId="17" LastEditorUserId="41" LastEditDate="2010-11-10T04:44:49.043" LastActivityDate="2015-10-06T21:04:10.933" Title="**How should I cook bacon in an oven?**" Tags="**&lt;oven&gt;&lt;cooking-time&gt;&lt;bacon&gt;**" AnswerCount="14" CommentCount="3" FavoriteCount="5" />

<row Id="3" PostTypeId="1" AcceptedAnswerId="25" CreationDate="2010-07-09T19:08:47.673" Score="65" ViewCount="6615" Body="**&lt;p&gt;I always use brown extra large eggs, but I can't honestly say why I do this other than habit at this point. Are there any distinct advantages or disadvantages like flavor, shelf life, etc?&lt;/p&gt;&#xA;**" OwnerUserId="14" LastEditorUserId="1672" LastEditDate="2013-03-24T21:42:44.057" LastActivityDate="2014-02-27T17:52:57.573" Title="**What is the difference between white and brown eggs?**" Tags="**&lt;eggs&gt;**" AnswerCount="4" CommentCount="1" FavoriteCount="2" />

<row Id="4" PostTypeId="1" AcceptedAnswerId="13" CreationDate="2010-07-09T19:09:20.653" Score="43" ViewCount="6550" Body**="&lt;p&gt;And can I use one in place of the other in certain recipes?&lt;/p&gt;&#xA;**" OwnerUserId="24" LastEditorUserId="1443" LastEditDate="2010-07-31T06:42:51.347" LastActivityDate="2010-07-31T06:42:51.347" Title="**What is the difference between baking soda and baking powder?**" Tags="**&lt;substitutions&gt;&lt;please-remove-this-tag&gt;&lt;baking-soda&gt;&lt;baking-powder&gt;**" AnswerCount="4" CommentCount="0" FavoriteCount="4" />

## Tags.xml

Enthält aller verwendeten Tags, sowie die Anzahl male, die sie verwendet wurden. Kann verwendet werden, um die relevanten Klassen im Multi-Label-Classifier Fall zu extrahieren:

<row Id="1" TagName="**cookies**" Count="303" ExcerptPostId="**20968**" WikiPostId="20967" />

<row Id="2" TagName="**baking**" Count="1767" ExcerptPostId="**13744**" WikiPostId="13743" />

<row Id="4" TagName="**bacon**" Count="80" ExcerptPostId="**23336**" WikiPostId="23335" />

<row Id="5" TagName="**eggs**" Count="510" ExcerptPostId="**6970**" WikiPostId="6969" />

<row Id="6" TagName="**please-remove-this-tag**" Count="**41**" ExcerptPostId="16128" WikiPostId="16127" />

## Badges.xml

Zuweizung von Badge zu User, nicht relevant:

<row Id="1" UserId="12" Name="Teacher" Date="2010-07-09T19:25:42.380" Class="3" TagBased="False" />

<row Id="2" UserId="17" Name="Teacher" Date="2010-07-09T19:25:42.550" Class="3" TagBased="False" />

<row Id="3" UserId="19" Name="Teacher" Date="2010-07-09T19:25:42.613" Class="3" TagBased="False" />

<row Id="4" UserId="25" Name="Teacher" Date="2010-07-09T19:25:42.660" Class="3" TagBased="False" />

## Comments.xml

Enthält alle Kommentare. Nicht relevant:

<row Id="1" PostId="6" Score="3" Text="Are you using fresh herbs or dry? What is the dish?" CreationDate="2010-07-09T19:11:09.683" UserId="19" />

<row Id="4" PostId="24" Score="22" Text="I make a whirpool in the water before dropping the eggs." CreationDate="2010-07-09T19:16:48.047" UserId="21" />

<row Id="9" PostId="29" Score="0" Text="Well, I'm using both brown and white sugar (as in every receipt I found), tried with baking soda and true leavening, but they become still crisp..." CreationDate="2010-07-09T19:21:25.230" UserId="16" />

<row Id="10" PostId="29" Score="0" Text="Replacing some of the white sugar with brown sugar would be where I would start." CreationDate="2010-07-09T19:21:46.460" UserId="50" />

## PostHistory.xml

Zeigt die Historie der Posts, sprich alle versionen wie sie im Laufe der Zeit vom Benutzer editiert wurden. Könnte anstatt der Posts.xml fürs Training verwendet werden (Man würde wohl die erste Version anstatt der letzten verwenden wollen):

<row Id="1" PostHistoryTypeId="2" PostId="1" RevisionGUID="e0907328-a66c-46b4-9bc5-f0328530c848" CreationDate="2010-07-09T19:07:45.460" UserId="16" Text="**My chocolate chips cookies are always too crisp. How can I get chewy cookies, like those of Starbucks?**" />

<row Id="2" PostHistoryTypeId="1" PostId="1" RevisionGUID="e0907328-a66c-46b4-9bc5-f0328530c848" CreationDate="2010-07-09T19:07:45.460" UserId="16" Text="**Chewy chocolate chips cookies**" />

<row Id="3" PostHistoryTypeId="3" PostId="1" RevisionGUID="e0907328-a66c-46b4-9bc5-f0328530c848" CreationDate="2010-07-09T19:07:45.460" UserId="16" Text="**&lt;cookies&gt;&lt;baking&gt;&lt;chocolate-chips-cookies&gt;**" />

Die Auswertung ist etwas complex, sodas man eher die Posts.xml-Datei verwenden wird.

## PostLinks.xml

Verlinkung der Posts (z.B. verweise zwischen Posts). Nicht relevant:

<row Id="66" CreationDate="2010-07-12T15:59:34.050" PostId="701" RelatedPostId="670" LinkTypeId="1" />

<row Id="120" CreationDate="2010-07-13T21:49:22.453" PostId="804" RelatedPostId="692" LinkTypeId="1" />

<row Id="149" CreationDate="2010-07-14T15:45:24.120" PostId="936" RelatedPostId="931" LinkTypeId="1" />

## Users.xml

Liste aller im Forum aktive Benutzer. Nicht relevant:

<row Id="5" Reputation="101" CreationDate="2010-07-09T17:38:25.903" DisplayName="**Kyle Cronin**" LastAccessDate="2015-11-08T20:43:14.713" AboutMe="&lt;p&gt;&lt;img src=&quot;http://media.podcast.askdifferent.net/tracking/profile.gif&quot; alt=&quot;&quot;&gt;&lt;/p&gt;&#xA;" Views="37" UpVotes="**1**" DownVotes="**0**" Age="31" AccountId="508" />

<row Id="6" Reputation="101" CreationDate="2010-07-09T18:42:29.710" DisplayName="**Kevin Montrose**" LastAccessDate="2017-06-27T03:12:13.437" WebsiteUrl="http://kevinmontrose.com/" Location="New York, NY, United States" AboutMe="&lt;p&gt;&lt;a href=&quot;http://blog.stackoverflow.com/2010/03/welcome-stack-overflow-valued-associate-000004/&quot;&gt;Stack Overflow Valued Associate #00004&lt;/a&gt;&lt;/p&gt;&#xA;&#xA;&lt;p&gt;For no good reason, &lt;a href=&quot;http://twitter.com/kevinmontrose&quot; rel=&quot;nofollow noreferrer&quot;&gt;I'm on twitter&lt;/a&gt;. I hear its compelling... somehow.&lt;br/&gt;&lt;/p&gt;&#xA;&#xA;&lt;p&gt;I also have a &lt;a href=&quot;http://kevinmontrose.com&quot; rel=&quot;nofollow noreferrer&quot;&gt;blog&lt;/a&gt;.&lt;/p&gt;&#xA;" Views="17" UpVotes="**39**" DownVotes="**0**" AccountId="**29738**" />

## Votes.xml

Upvotes und Downvotes. Nicht relevant:

<row Id="1" PostId="1" VoteTypeId="2" CreationDate="2010-07-09T00:00:00.000" />

<row Id="2" PostId="1" VoteTypeId="2" CreationDate="2010-07-09T00:00:00.000" />

<row Id="3" PostId="2" VoteTypeId="2" CreationDate="2010-07-09T00:00:00.000" />

<row Id="4" PostId="1" VoteTypeId="2" CreationDate="2010-07-09T00:00:00.000" />

<row Id="5" PostId="3" VoteTypeId="2" CreationDate="2010-07-09T00:00:00.000" />

# Trainingsdaten

Die Daten werden vom IBM Watson Natural Language Classifier als CSV-Datei erwartet. Dabei wird pro Datensatz maximal 1024 Charakter zugelassen, und maximal 20'000 Datensätze insgesamt sind erlaubt (zum Vergleich: Allein cooking.stackexchange.com hat über 65000 Posts).

Aus den oben genannten Beispiele würde sich für den Single-Label-Classifier potentiell folgende Datei ergeben:

**“How can I get chewy chocolate chip cookies? \r My chocolate chips cookies are always too crisp. How can I get chewy cookies, like those of Starbucks? Thank you to everyone who has answered. So far the tip that had the biggest impact was to chill and rest the dough, however I also increased the brown sugar ratio and increased a bit the butter. Also adding maple syrup helped.”, cooking.stackexchange.com**

**“How should I cook bacon in an oven? \r I've heard of people cooking bacon in an oven by laying the strips out on a cookie sheet. When using this method, how long should I cook the bacon for, and at what temperature?”,** **cooking.stackexchange.com**

**“What is the difference between white and brown eggs? \r I always use brown extra large eggs, but I can't honestly say why I do this other than habit at this point. Are there any distinct advantages or disadvantages like flavor, shelf life, etc?”,** **cooking.stackexchange.com**

**“What is the difference between baking soda and baking powder? And can I use one in place of the other in certain recipes?”,** **cooking.stackexchange.com**

Für den Multi-Label-Classifiert beispielsweise diese:

**“How can I get chewy chocolate chip cookies? \r My chocolate chips cookies are always too crisp. How can I get chewy cookies, like those of Starbucks? Thank you to everyone who has answered. So far the tip that had the biggest impact was to chill and rest the dough, however I also increased the brown sugar ratio and increased a bit the butter. Also adding maple syrup helped.”, baking, cookies, texture**

**“How should I cook bacon in an oven? \r I've heard of people cooking bacon in an oven by laying the strips out on a cookie sheet. When using this method, how long should I cook the bacon for, and at what temperature?”,** **oven, cooking-time, bacon**

**“What is the difference between white and brown eggs? \r I always use brown extra large eggs, but I can't honestly say why I do this other than habit at this point. Are there any distinct advantages or disadvantages like flavor, shelf life, etc?”,** **eggs**

**“What is the difference between baking soda and baking powder? And can I use one in place of the other in certain recipes?”,** **substitutions, baking-soda, baking-powder**

<https://console.bluemix.net/docs/services/natural-language-classifier/using-your-data.html#data-preparation>