

فصل دوم

Lemma: پیداکردن ریشه کلمه. مثلاً: می‌نویسم، نوشتند ← رفتن

stemming: پیدا کردن بن که با حذف و اضافه اضافات همراه است. مثلاً: می‌نویسم، نوشتند ← رفت

wordform: شکل کل صفت یک کلمه. stem یک نسبت از صفت [وزن] که بیان کلمه [stems] اضافه کرده اند.

نمایش داده می‌شود. هر دو هم + رو + شم

Type: یک فصل از دایره نام [هیچ دو مدخلی یکسان نیستند] مثلاً در یک متن they و their دلالت یکسان

Token: یک نمونه از یک فصل که در متن استفاده شده است. هستند ولی توکن کل میز به حساب می‌آید

سوال ۱۴.

الف. Type ب. Token ج. Wordform د. Lemma

تعاریف:

$$accuracy = \frac{tp + tn}{tp + tn + fp + fn}$$

یعنی در بین تمام حالات، چند بار تلاش درست تشخیص داده شده است؟

$$precision = \frac{tp}{tp + fp}$$

یعنی در بین تمام حالات که می‌گوئیم تلاش انتخاب شده است، چند بار تلاش درست انتخاب شده است؟

$$recall = \frac{tp}{tp + fn}$$

یعنی در بین تمام حالات که درست است، چند بار آن تلاش درست انتخاب شده است؟

	correct	incorrect	
selected	tp	fp	precision
not selected	fn	tn	accuracy
			recall

سوال ۱۵.

فرض: هر کلمات مورد انتظار را می‌تواند در مجموعه بر آن که، کلمات که شرایط مورد انتظار را دارند هم بر می‌تواند.

جواب: معیار recall.

باقیه به ایند از بین حالات که درست و مورد انتظار، تمامی درست که را تشخیص می‌دهد یعنی در این مسئله معیار recall برابر 100 خواهد بود. [بهر طریقی که باید به عنوان درست انتخاب می‌شود، حتماً برپایه نوشته است پس: هیچ کلمه‌ای درستی که به اشتباه انتخاب نشده باشد در این مجموعه نیست پس $fn = 0$ است پس $recall = 1$ خواهد بود]

نکته اضافه: باید دقت داشت که از توکل در معیار recall عملکرد خوبی را ارزیابی می‌کنند ولی از توکل در معیار این عبارت مناسب تر است معیار precision مناسب تر است. چون که recall، جواب کمی که به اشتباه، درست تشخیص داده شده از آن [fp] در نظر گرفته و درین قوم به آنها، معیار 1 را گزارش می‌کند

سؤال 16.

f_n

فصل دوم

فرض: عبارت منظم می تواند نامرئی باشد و اگر به دنبال آن بوده ایم نیز باید در آن سرطانی بیرون باشد و آن کلمه همان کلمه است که مورد انتظار را بگوید.

جواب: Precision، Recall

* با توجه به اینکه عبارت منظم داده شده از بین تمام حالات با یکدیگر یک صلاصت دارند و این صلاصت در لغت نیز یک کلمه و به عنوان کلمه درست بر می آید و از آنجا که تمام کلمات بیرون داده شده کلمه درست است پس هیچ کلمه بیرون داده نمی شود که غلط بوده باشد و عبارت منظم آنرا را از آنها درست تشخیص داده باشد [FP=0].
بنابراین معیار Precision بهتر از Recall خواهد بود و ممکن است که از این عبارت منظم را ارزیابی کند.

نکته اضافه:

← البته باید توجه کرد که اگر بخواهیم درست تشخیص داده می شوند به اشتباه، نادرست تشخیص داده شده و به عنوان کلمه صحیح از عبارت منظم بیرون داده می شود از این معیار Precision علاوه بر این صحت و به صورت کاذب معیار 1 را تراش می کند در حالی که معیار Recall برای عبارت منظم ارزیابی خوبی خواهد داشت.

سؤال 17.

با توجه به اینکه شکل نوشتاری کلمات و کلمات با هم شباهت دارند و اینها با مفاد خاص همراه است و طبعاً یک اسم خاص است که با تغییر حرف نزدیک به حرف دیگر، یا معنای مورد انتظار تغییر کرده و در حال اسم خاص اولیه مورد هدف به عنوان یک کلمه عام توسط ماشین ترجمه به زبان مقصد ترجمه شود. بنابراین Case folding در اینجای کاربرد است.

مثلاً "General Motors" درست ترجمه به فارسی "جنرال موتورز" و "general motors" در صورت ترجمه به فارسی، مورد کاربرد معمولی می تواند ترجمه شود در حالی که Case folding باعث بروز خطا در خروجی ترجمه شده است.

s.am