

دانشگاه صنعتی شریف

دانشکده مهندسی کامپیوتر



بازیابی پیشرفته اطلاعات

تمرین شماره دو (فردی)

پردازش متن

امیرحسین باقری ۹۸۱۰۵۶۲۱

استاد:

دکتر عسکری

فهرست مطالب

۲	۱	جمع آوری داده ها
۲	۱.۱	کراال کردن سایت گنجور
۲	۲	پیش پردازش داده های شاعران
۲	۱.۲	کراال کردن قرن هر شاعر
۳	۲.۲	مهیا نمودن نوشته های هر قرن
۳	۳	بررسی محتوای ادبیات پارسی از قرن ۳ تا کنون
۳	۱.۳	مراحل پردازش بر روی داده های قرن
۴	۲.۳	بررسی ترکیبات اضافی و ترکیبات وصفی و لغات
۴	۴	نتایج بررسی دیتا های قرن ها و تحلیل نتایج
۵	۱.۰.۴	قرن ۳م تا ۱۴ در یک نگاه
۶	۲.۰.۴	سیر کلمات و ترکیبات در طی قرون ۳ تا ۱۴
۸	۳.۰.۴	بررسی correlation مفاهیم در ادبیات فارسی (بر حسب درصد)
۱۰	۴.۰.۴	نمودار مربوط به تمامی قرن ها ۳ تا ۱۴
۲۱	۵	۵ بررسی ۶ شاعر پرمخاطب سایت گنجور و مقایسه آنها با یکدیگر با استفاده از دیتا های بدست آمده از قرن های ۳ تا ۱۴
۲۲	۱.۵	ترکیبات و کلمات پر تکرار
۲۳	۱.۱.۵	سعدی
۲۵	۲.۱.۵	حافظ
۲۷	۳.۱.۵	مولوی
۲۹	۴.۱.۵	خیام
۳۱	۵.۱.۵	نظامی
۳۳	۶.۱.۵	فردوسی

مقدمه

کاری که در پی انجام آن هستیم بررسی محتوای شاعران معروف در سایت گنجور و مقایسه آنها با محتوای ادبیات فارسی در تاریخ است. این کار در ۴ فاز انجام شده است.

۱. جمع آوری نوشته های تمام شعرا و نویسندگان سایت گنجور (حدوداً ۱۸۰ مگابایت داده جمع شده است).

۲. پردازش داده های جمع شده و آماده سازی متن های قرن های ۳ ۵ ۶ ۷ ۸ ۹ ۱۰ ۱۱ ۱۲ ۱۳ ۱۴ برای بررسی آثار.

۳. پردازش متن های تمیز شده برای بررسی محتوای آنها

۴. پردازش داده های حافظ، سعدی، فردوسی، خیام، مولوی و نظامی که ۶ شاعر پرمخاطب در سایت گنجور هستند.

دقت کنید که دیگر به توضیح کدها نمی پردازیم و تنها به مراحل انجام شدن کار می پردازیم.

۱ جمع آوری داده ها

۱.۱ کرال کردن سایت گنجور

کرال این سایت به وسیله فریم ورک scrapy انجام شده است. ابتدا تمام شاعران و صفحه های آن انجام شده است برای این کار کافیت کد زیر را در پوشه data crawler اجرا کنید. (درون شل و دقت کنید باید فریم ورک scrapy را نصب کنید).

```
pip install Scrapy
scrapy crawl all_poets
```

بدین صورت کرال تمام شاعران به همراه لینک صفحه هایشان انجام می شود این کرال حدوداً ۵ دقیقه زمان میبرد.

کرال این سایت به شیوه DFS صورت گرفته بدان معنا که تمامی لینک های سایت جمع آوری و سپس همان لینک ها کرال می شود تا به صفحه آخر که در آخر لینک آن یک رجکس خاص قرار دارد برسیم. سپس متن های مربوطه درون آن صفحه کرال می شود. برای دیدن کدهای کرال به پوشه زیر می توانید برید.

```
data_crawler / data_crawler / spiders / all_poets_name_links.py
```

پس از جمع آوری لینک های صفحه ها اسکریپت زیر را اجرا میکنید.

```
cd data_crawler
python3 crawl_all_script.py
```

این کرال ۶ ساعت به طول می انجامد. و محتوای هر شاعر داخل پوشه به نام آن شاعر قرار میگیرد.

```
data/<poet name>
data/<poet name>/<poet name.txt>
```

۲ پیش پردازش داده های شاعران

۱.۲ کرال کردن قرن هر شاعر

ابتدا شاعران هر قرن را کرال میکنیم برای اینکار اسکریپت زیر را در ترمینال اجرا کنید.

```
cd data_crawler
scrapy crawl all_poets_c
```

۲.۲ مهیا نمودن نوشته های هر قرن

حال نوتبوک HW2_data_preprocess.ipynb را اجرا میکنیم در این نوتبوک داده های هر قرن را بر حسب شاعر استخراج کرده و در فایل به نام عدد آن قرن ذخیره میکنیم. پس از این کار شما باید درون پوشه century_data بتوانید فایل های تکس مربوط به هر قرن را مشاهده کنید.

۳ بررسی محتوای ادبیات پارسی از قرن ۳ تا کنون

۱.۳ مراحل پردازش بر روی داده های قرن

در نوتبوک HW2_century.ipynb مراحل زیر اجرا می شوند.

۱. ابتدا هر جمله را با کمک نرمالایزر نرمالایز میکنیم.

۲. سپس با کمک tokenizer کلمات هر جمله را جدا میکنیم.

۳. سپس با کمک lemmatizer هر کلمه را به ریشه آن برمی گردانیم. دقت کنید که در این مرحله اگر tagger واژه را فعل یا حرف اضافه تشخیص دهد آنرا حذف میکنیم. واژه هایی که داخل stopwords باشند نیز حذف می شوند.

۴. فریکوئنسی داده ها را بررسی میکنیم. سپس توسط کلمات که بیشترین تکرار را داشته اند لغات زائد را ابدیت می کنیم.

```
module_data/my_stopwords.txt
```

۶۰۰ کلمه و حرف جدید به لغات stopwords اضافه می کنیم. (دقت کنید که در ابتدای لود کردن نوت بوک لغات زائد را با لغات زائدی که قبلا داشته ایم مرج میکنیم فلذا با یکبار ران کردن دوباره نوتبوک فهرست لغات زاید ابدیت شده اند.)

۵. ۲۵۰ لغت پر تکرار هر قرن را در یک فایل اکسل ذخیره می کنیم. که به ۷۰۰ لغت یونیک ختم می شود. سپس این لغات رو بررسی میکنیم.

۶. همچنین با کمک

```
def vaf_ext(sentence):
    grammar = r"""
    NVASF: <N|Ne><AJ>
    NEZAF: <N|Ne><N|Ne>
    """
    cp = nltk.RegexpParser(grammar)
    return (cp.parse(sentence))
    tarkibat = {}
for key in word_tokenized:
    tarkibat[key] = {}
    NVASF = []
    NEZAF = []
    for sentence in tqdm.tqdm(word_tokenized[key]):
        tree=vaf_ext(tagger.tag(sentence))
```

```

for subtree in tree.subtrees():
    if subtree.label() == 'NVSF':
        NVASF.append(subtree.leaves())
    if subtree.label() == 'NEZAF':
        NEZAF.append(subtree.leaves())
tarkibat[key]["NVSF"] = ['_'].join([x[0],y[0]]) for x,y in NVASF
tarkibat[key]["NEZAF"] =['_'].join([x[0],y[0]]) for x,y in NEZAF]

```

۷. ترکیبات وصفی و ترکیبات اضافی ممتن را استخراج میکنیم (۲۰۰ ترکیب برتر هر قرن) به ۱۲۰۰ ترکیب اضافی و وصفی ختم می‌شود.

۲.۳ بررسی ترکیبات اضافی و ترکیبات وصفی و لغات

این ترکیبات را به صورت زیر نشان می‌دهیم.
 تمامی ترکیبات اضافی با پسوند NEZAF مشخص می‌شوند.
 تمامی ترکیبات وصفی با پسوند NVASF مشخص می‌شوند.
 لغت‌ها با پسوند word_ مشخص می‌شوند.
 سپس بررسی می‌کنیم که این لغات و ترکیبات از کدام یک از دسته‌های زیر هستند.

۱. love: لغات و ترکیباتی که بار عارفانه و عاشقانه دارند.

۲. philosophy: لغات و ترکیباتی که بار فلسفی و توصیه دارند. و یا به مسایل اخلاقی و سیاسی می‌پردازند.

۳. religion: لغات و ترکیباتی که بار مذهبی و دینی دارند.

۴. daily: لغاتی که به مسایل عادی و روزمره مانند سبزه زار و آسمان و خوشی و غم‌های دنیایی می‌پردازند.

۵. others: لغاتی که در ۴ دسته بندی بالا قرار نمی‌گیرند.

هر کدام از این دسته بندی‌ها به صورت دستی صورت گرفته اند. دقت کنید می‌توانستیم از دیتاست فارسی نت استفاده کنیم اما محدودیت ۱۰۰۰ کلمه در هفته مانع این کار می‌شود. این دیتاست دسته بندی شده در پوشه زیر قرار دارند. type_dataset

```

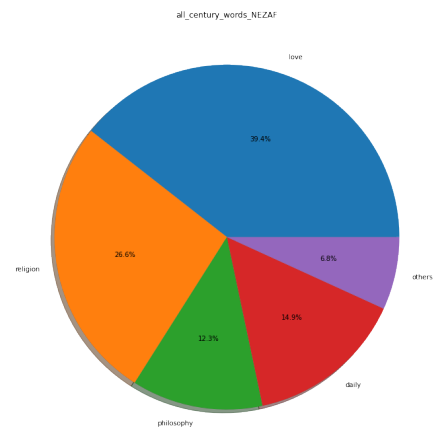
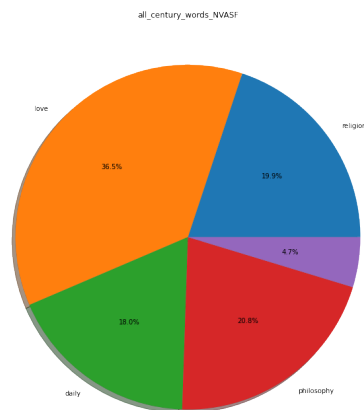
type_dataset/unique_most_words
type_dataset/unique_most_NVSF
type_dataset/unique_most_NEZAF

```

۴ نتایج بررسی دیتاهای قرن‌ها و تحلیل نتایج

تمامی فایل‌ها با پسوند های مرتبط مشخص شده اند که مرتبط به چه قرنی هستند.
 همچنین در همان پسوند ها مشخص شده اند که به ترکیبات اضافه مرتبط اند یا وصفی یا لغات
 ۲۵۰ لغت پرتکرار بررسی شده اند و نتایج آنها در زیر آمده است.

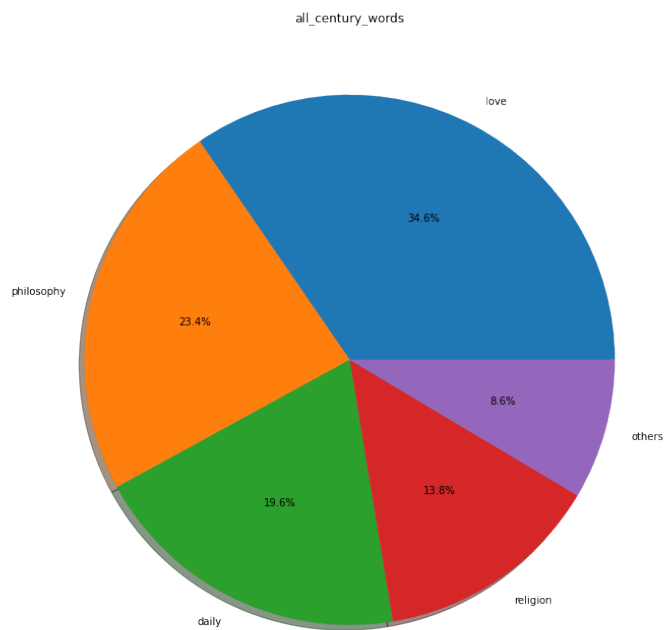
۱.۰.۴ قرن ۳ام تا ۱۴ در یک نگاه



شکل ۱: ترکیبات اضافی

شکل ۲: ترکیبات وصفی

۵.



شکل ۳: کلمات

شکل ۴: ادبیات فارسی در یک نگاه

14	13	12	11	10	9	8	7	6	5	3
(خون دل, 92)	(سر زلف, 166)	(خون دل, 91)	(بوی گل, 311)	(آب حیات, 117)	(غم عشق, 100)	(بسمت الله, 675)	(آب حیات, 245)	(رسول خدا, 1009)	(شیخ الاسلام, 800)	(فاخته گون, 4)
(خون جگر, 54)	(روز قیامت, 161)	(خون جگر, 59)	(نقش پا, 253)	(خون دل, 105)	(این حسام, 97)	(این بمن, 489)	(سر زلف, 174)	(آب این, 466)	(حق تعالی, 265)	(گردش خورشید, 3)
(آب صبا, 42)	(حضرت امام, 156)	(بوی گل, 55)	(خون دل, 245)	(ای موش, 102)	(خون دل, 86)	(سر زلف, 419)	(خون دل, 171)	(این عیان, 459)	(استاد ابو طیب, 158)	(سر ای سینج, 3)
(سر زلف, 34)	(خون دل, 148)	(کام دل, 50)	(شاخ گل, 214)	(خون جگر, 95)	(سر کوی, 85)	(سر کوی, 341)	(شمس تبریزی, 163)	(رسول الله, 441)	(ترجمه الله, 141)	(نام نیکم, 2)
(علی جان, 33)	(سلام مروی, 127)	(شاخ گل, 43)	(خون جگر, 204)	(برد دل, 93)	(خون جگر, 75)	(سیف فرغانی, 333)	(سر کوی, 151)	(بیر طریقت, 309)	(شاه جهان, 133)	(عاشقان نشین, 2)

شکل ۵: ترکیبات اضافی

14	13	12	11	10	9	8	7	6	5	3
(آه دور, 91)	(خدای تعالی, 184)	(سور رون, 23)	(آب رون, 201)	(آب پر, 62)	(مه نو, 52)	(آب خسته, 131)	(چشم بد, 134)	(خدای تعالی, 361)	(خدای تعالی, 1301)	(گل سرخ, 6)
(رنج بسوز, 81)	(ماه نو, 59)	(آب تنگ, 21)	(ماه نو, 200)	(مه نو, 56)	(عالم پر, 30)	(آب رون, 117)	(آب رون, 83)	(الله تعالی, 241)	(حق تعالی, 301)	(آهوان پر, 3)
(خواب گران, 43)	(آب پر, 52)	(خواب گران, 20)	(آریگ رون, 172)	(ماه نو, 45)	(ماه نو, 30)	(سور رون, 106)	(سور رون, 78)	(آب بکر, 167)	(خدای عزوجل, 193)	(عشاق کم, 2)
(شیر نر, 38)	(آب رون, 50)	(آب رون, 18)	(بولت بیدار, 164)	(بخت بد, 40)	(چشم بد, 27)	(شب نیرم, 101)	(حق تعالی, 61)	(حق جل, 163)	(خداوند تعالی, 185)	(باغ خرم, 2)
(گل سرخ, 30)	(خند برین, 49)	(چشم سیم, 18)	(خواب گران, 161)	(خداوند تعالی, 36)	(دل تنگ, 27)	(کرز گران, 97)	(آب تنگ, 56)	(خدای عز, 155)	(نفس کلی, 180)	(نبیذ روشن, 2)

شکل ۶: ترکیبات وصفی

14	13	12	11	10	9	8	7	6	5	3
(جان, 2528)	(جان, 6523)	(عشق, 1955)	(چشم, 12392)	(جان, 4635)	(جان, 4442)	(جان, 12186)	(جان, 15149)	(الله, 29867)	(جان, 6932)	(جهان, 49)
(نست, 2370)	(عشق, 5128)	(جان, 1890)	(نست, 9477)	(عشق, 3768)	(عشق, 3820)	(عشق, 7790)	(عشق, 7606)	(جان, 18073)	(جهان, 6858)	(جان, 38)
(کار, 2048)	(نست, 5103)	(چشم, 1741)	(آب, 9423)	(غم, 3672)	(چشم, 2896)	(چشم, 6751)	(نست, 7005)	(علی, 11719)	(نست, 6797)	(آباد, 38)
(چشم, 1851)	(چشم, 4409)	(نست, 1660)	(عشق, 7852)	(چشم, 3361)	(چشم, 2417)	(جهان, 6733)	(چشم, 6901)	(نست, 9891)	(آباد, 6602)	(چشم, 32)
(جهان, 1757)	(شاه, 3764)	(غم, 1412)	(صائب, 6972)	(نست, 3352)	(غم, 2376)	(نست, 6074)	(آب, 5568)	(جهان, 9529)	(شاه, 5895)	(آب, 31)

شکل ۷: کلمات

شکل ۸: ادبیات فارسی در یک نگاه

۲۰۰۴ سیر کلمات و ترکیبات در طی قرون ۳ تا ۱۴

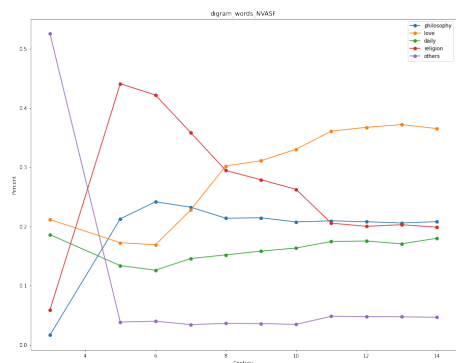
همانطور که در شکل ۱۱ مشاهده می شود. کلمات عاشقانه در طی قرن ۳م تا ۱۴م ابتدا سیر نزولی و سپس صعودی داشته اند و از قرن ۷م بیشترین درصد را به خودشان اختصاص داده اند.

کلمات دینی و مذهبی در قرن ۶م به اوج خودشان رسیدند.

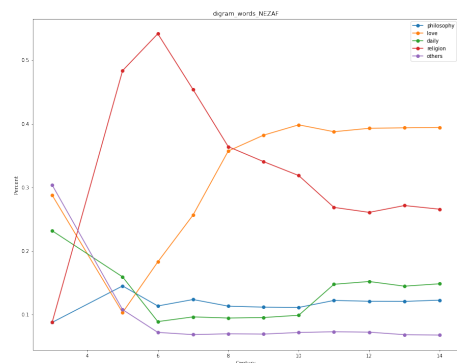
کلمات فلسفی در قرن ۵م بیشترین درصد را به خودشان اختصاص داده اند. کلمات زندگی روزمره سیر نزولی داشته اند.

در شکل ۱۰ مشاهده می شود که روند صعودی یا نزولی بودن ترکیبات وصفی با کلمات اندکی متفاوت است و مفاهیم دینی تا قرن ۸م دست بالا دارند. همچنین همانطور که میبینید روند مفاهیم عاشقانه مشابه کلمات است. بقیه مفاهیم نیز سیر مشابه ای را داشته اند.

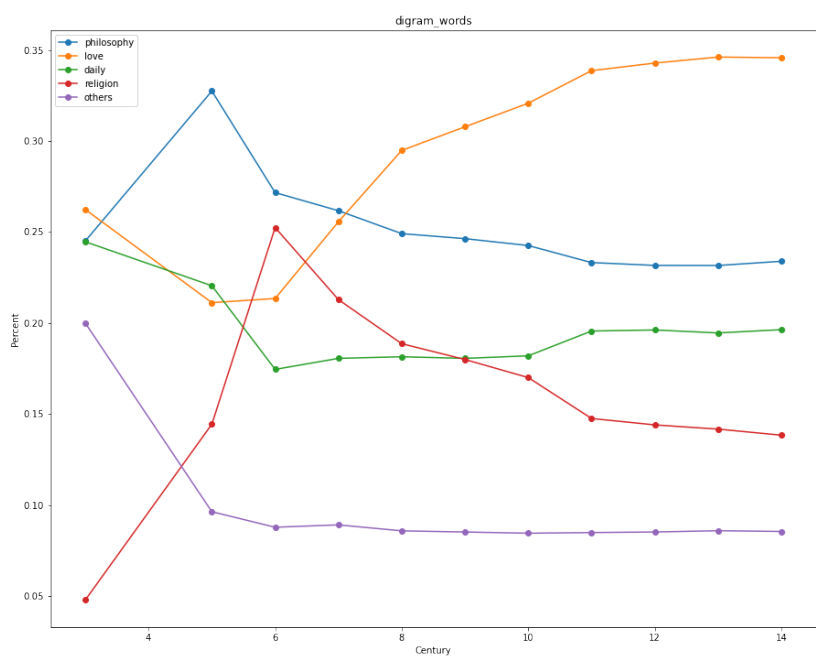
در شکل ۹ مشاهده می کنید که ترکیبات اضافی و وصفی سیر مشابه ای داشته اند.



شکل ۱۰: ترکیبات وصفی



شکل ۹: ترکیبات اضافی

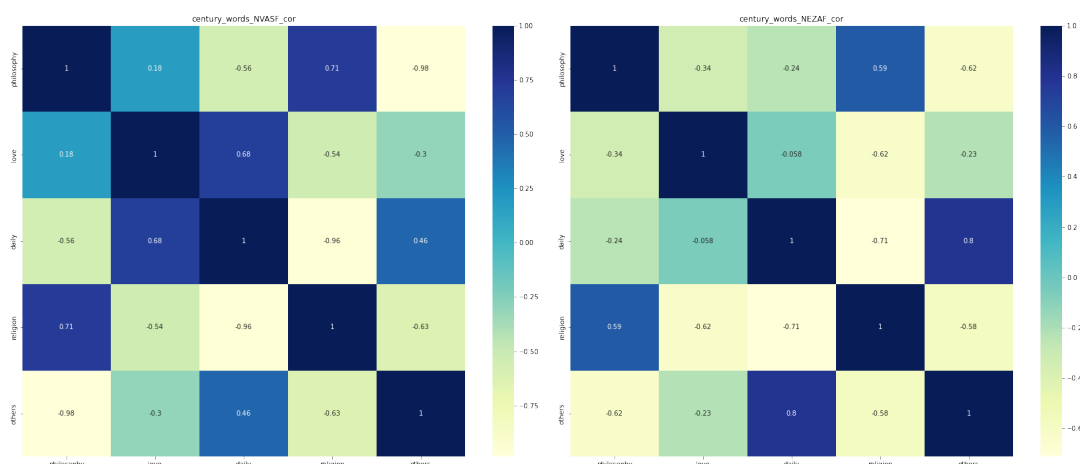


۵.

شکل ۱۱: کلمات

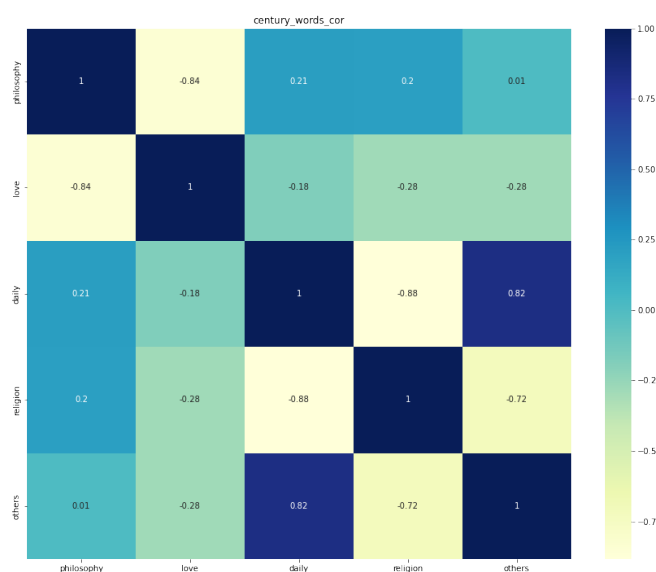
شکل ۱۲: روند تحول ادبیات فارسی در طی قرون ۳ تا ۱۴

۳.۰.۴ بررسی correlation مفاهیم در ادبیات فارسی (بر حسب درصد)



شکل ۱۴: ترکیبات وصفی

شکل ۱۳: ترکیبات اضافی



۵.

شکل ۱۵: کلمات

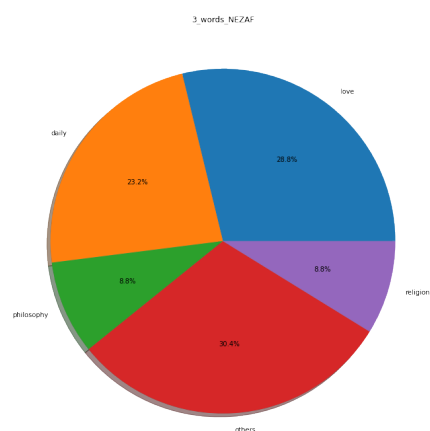
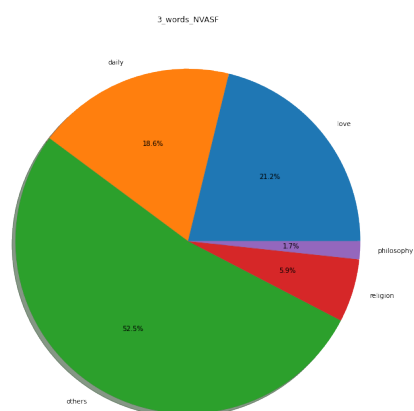
شکل ۱۶: همبستگی مفاهیم و ترکیبات در ادبیات فارسی

همانطور که در نمودارهای بالا مشاهده می کنید. کلمات فلسفی و عاشقانه همبستگی منفی دارند بدان معنا که با افزایش کلمات فلسفی کلمات عاشقانه کاهش می یابند. از نمونه های دیگر در این نمودار می توان به همبستگی منفی مفاهیم مذهبی و عارفانه اشاره کرد که در ادبیات فارسی آنها به تقابل شریعت و طریقت ارجاع می دهیم. مفاهیمی که به زندگی روزمره اشاره دارند همبستگی منفی با مفاهیم مذهبی دارند بدان معنا که پرداختن به مسائل دینی کاهش پرداختن به مفاهیم معمولی و روزانه را در پی داشته اند. در ترکیبات وصفی همبستگی منفی بیشتری بین مفاهیم عرفانی و عارفانه نسبت به کلمات دیده می شود.

و مفاهیم مذهبی و روزمره در این بخش همبستگی نزدیک به منفی یک دارند که تقریباً به معنای آن است که رابطه خطی دارند و با مثلاً ۲ برابر شدن یکی دیگری نصف می شود. در ترکیبات اضافی همبستگی منفی بین مذهب و عرفان بیشتر دیده می شود ولی مفاهیم مذهبی و فلسفی همبستگی نزدیک به ۰.۶ دارند که عدد قابل توجه ای است.

۴.۰.۴ نمودار مربوط به تمامی قرن ها ۳ تا ۱۴

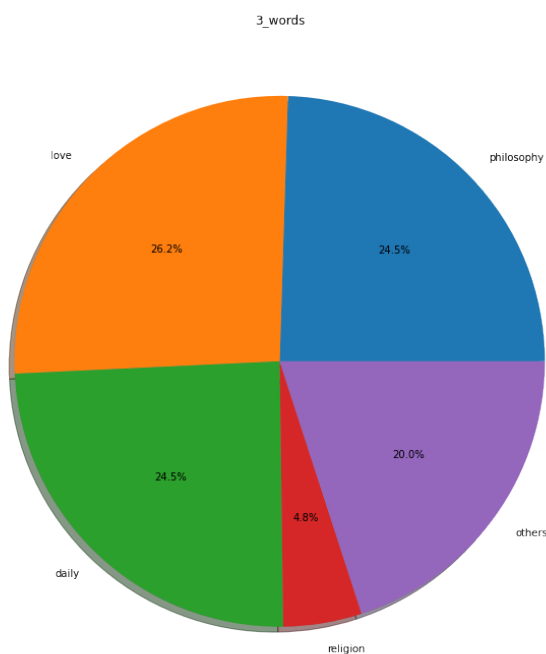
این نمودار ها سهم هر مفهوم را نشان می دهند و دیگر توضیح برایشان نمیدهیم.



شکل ۱۸: ترکیبات وصفی

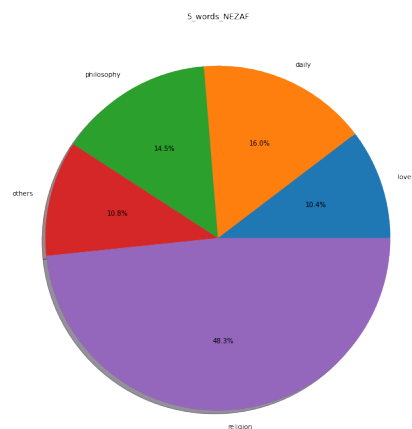
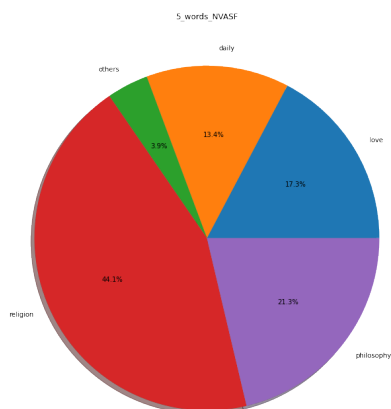
شکل ۱۷: ترکیبات اضافی

۵.



شکل ۱۹: کلمات

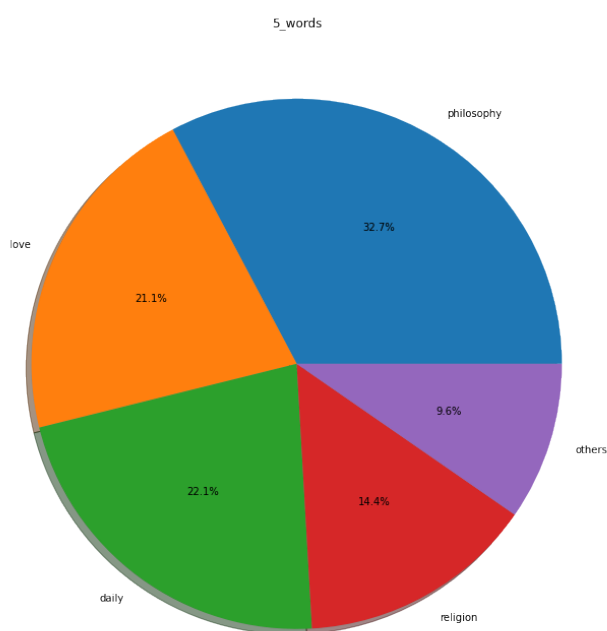
شکل ۲۰: سهم هر مفهوم از کلمات و ترکیبات وصفی و اضافی قرن ۳ ام



شکل ۲۱: ترکیبات اضافی

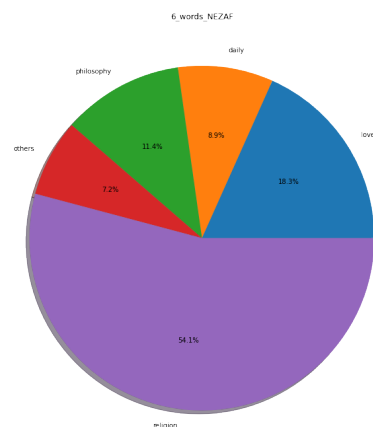
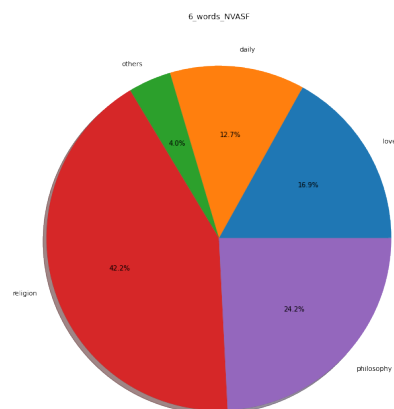
شکل ۲۲: ترکیبات وصفی

۵.



شکل ۲۳: کلمات

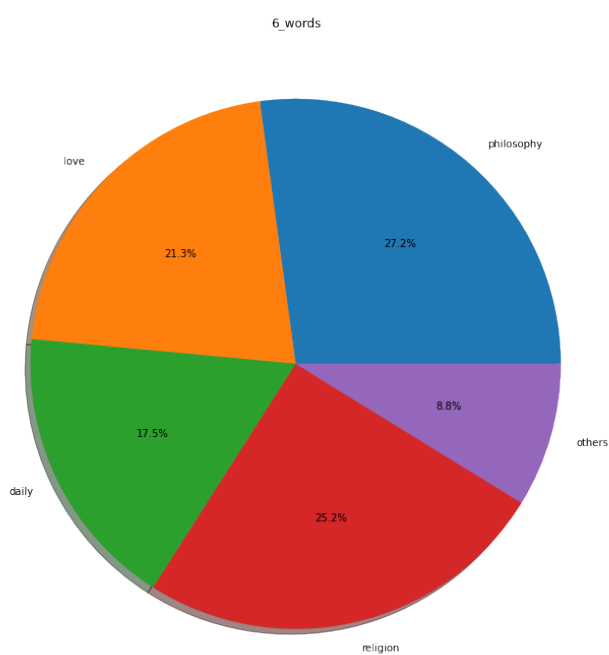
شکل ۲۴: سهم هر مفهوم از کلمات و ترکیبات وصفی و اضافی قرن ۵ ام



شکل ۲۵: ترکیبات اضافی

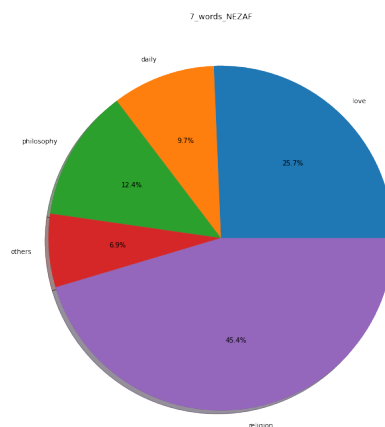
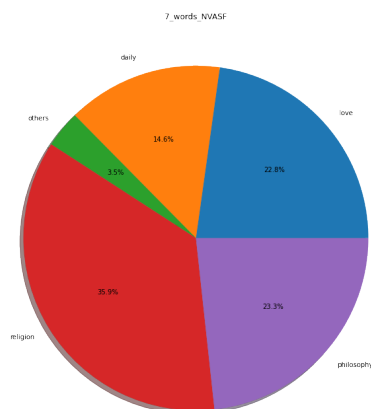
شکل ۲۶: ترکیبات وصفی

۵.



شکل ۲۷: کلمات

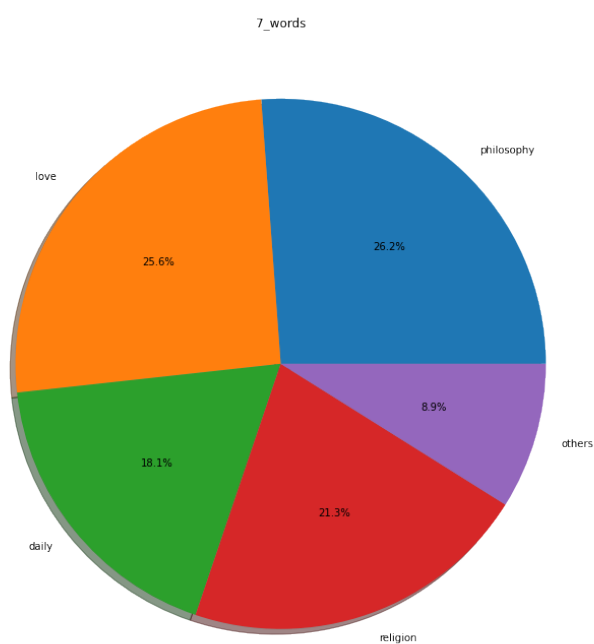
شکل ۲۸: سهم هر مفهوم از کلمات و ترکیبات وصفی و اضافی قرن ۱۶م



شکل ۲۹: ترکیبات اضافی

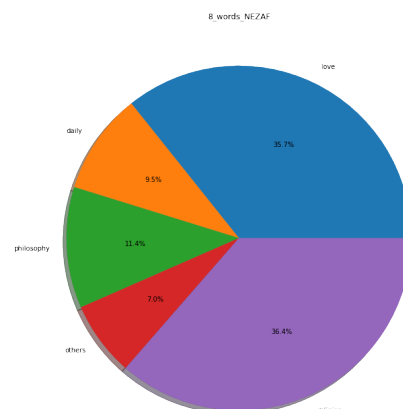
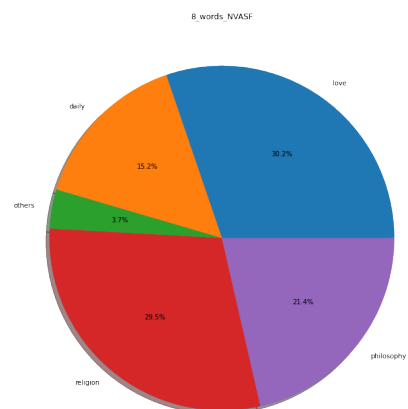
شکل ۳۰: ترکیبات وصفی

۵.



شکل ۳۱: کلمات

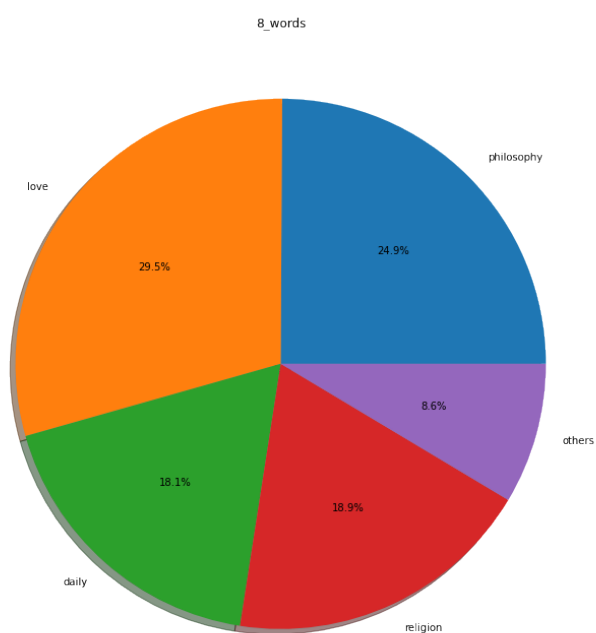
شکل ۳۲: سهم هر مفهوم از کلمات و ترکیبات وصفی و اضافی قرن ۱۷م



شکل ۳۳: ترکیبات اضافی

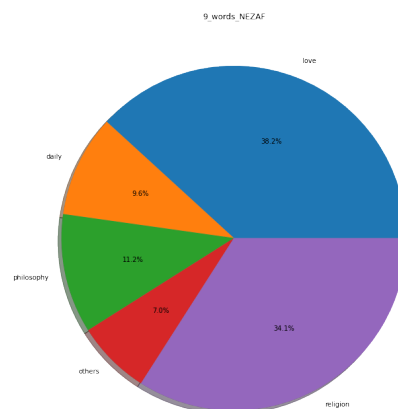
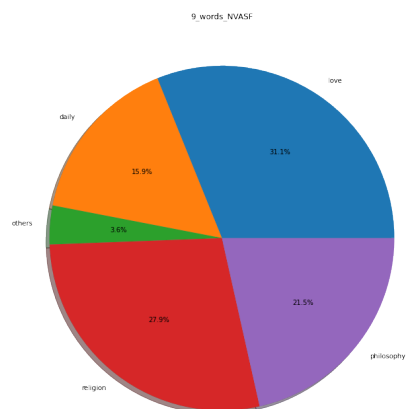
شکل ۳۴: ترکیبات وصفی

۵.



شکل ۳۵: کلمات

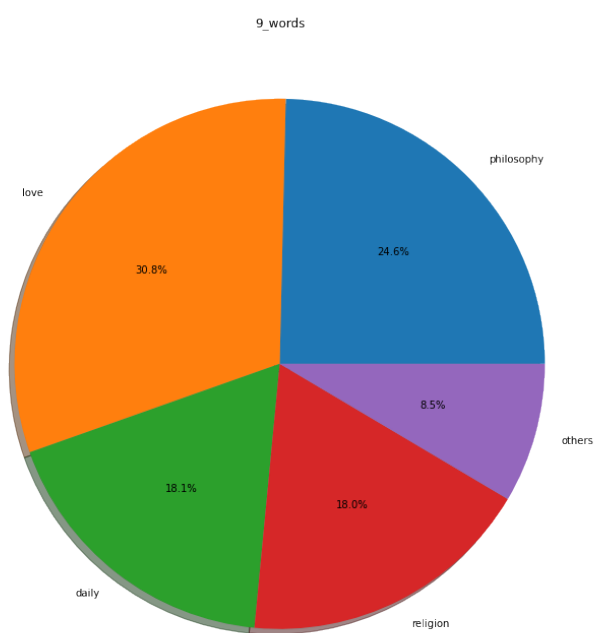
شکل ۳۶: سهم هر مفهوم از کلمات و ترکیبات وصفی و اضافی قرن ۱۸م



شکل ۳۷: ترکیبات اضافی

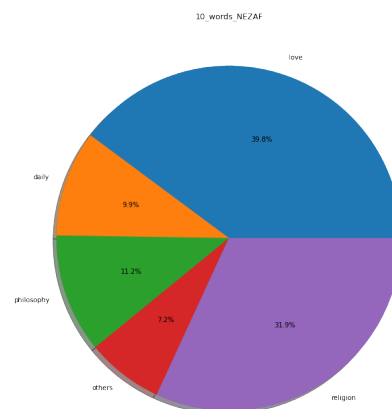
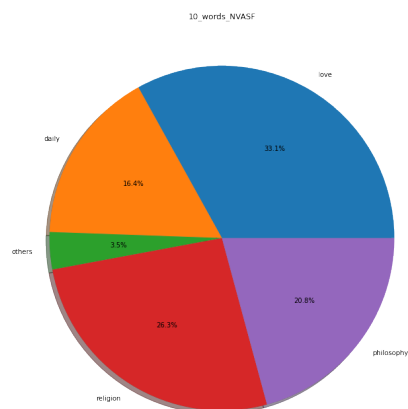
شکل ۳۸: ترکیبات وصفی

۵.



شکل ۳۹: کلمات

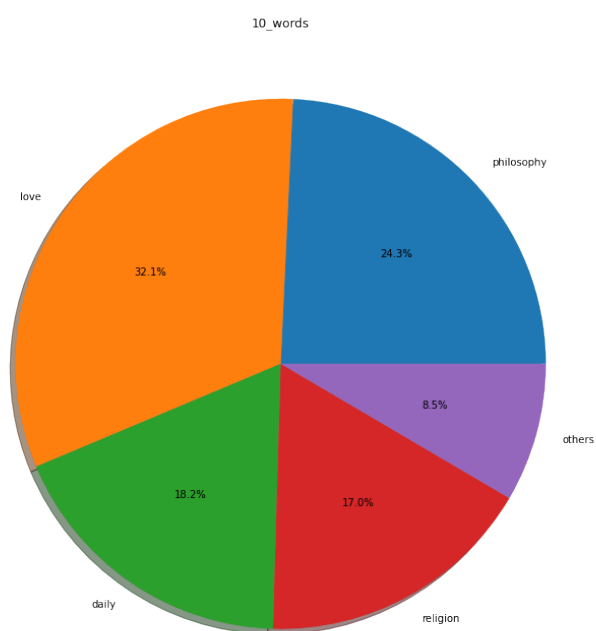
شکل ۴۰: سهم هر مفهوم از کلمات و ترکیبات وصفی و اضافی قرن ۹م



شکل ۴۱: ترکیبات اضافی

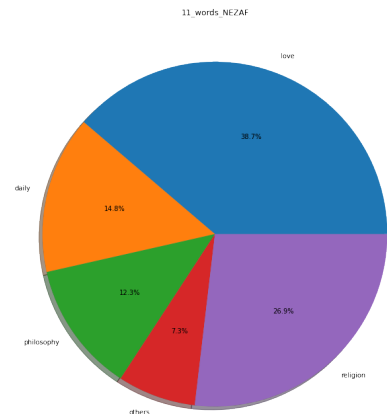
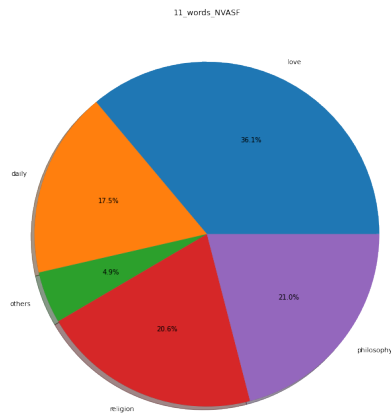
شکل ۴۲: ترکیبات وصفی

۵.



شکل ۴۳: کلمات

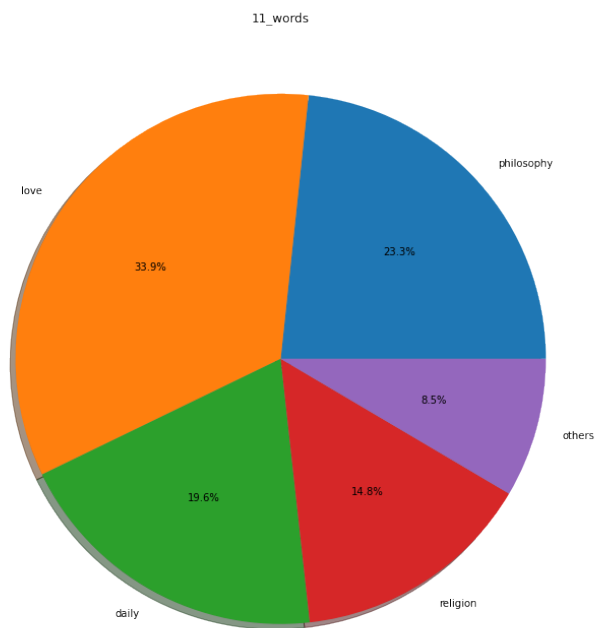
شکل ۴۴: سهم هر مفهوم از کلمات و ترکیبات وصفی و اضافی قرن ۱۰ ام



شکل ۴۵: ترکیبات اضافی

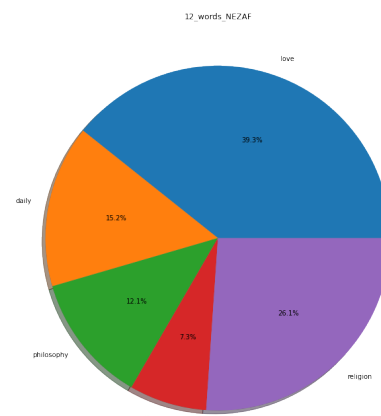
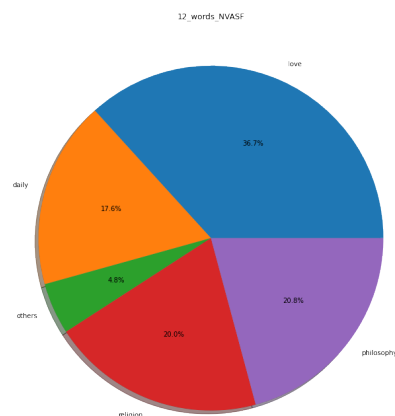
شکل ۴۶: ترکیبات وصفی

۵.



شکل ۴۷: کلمات

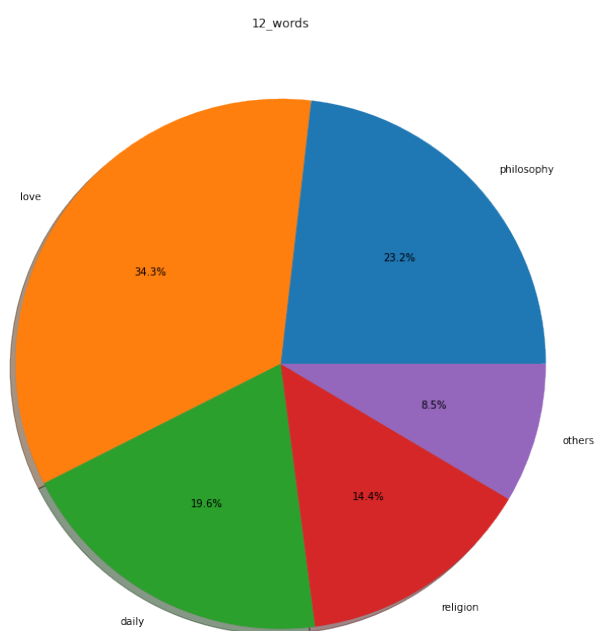
شکل ۴۸: سهم هر مفهوم از کلمات و ترکیبات وصفی و اضافی قرن ۱۱ ام



شکل ۴۹: ترکیبات اضافی

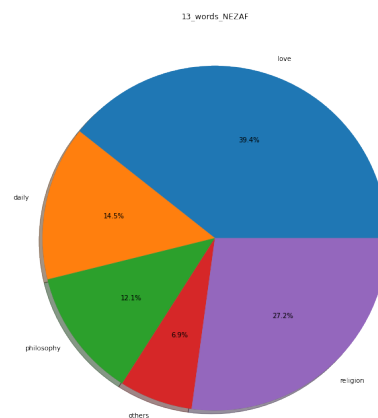
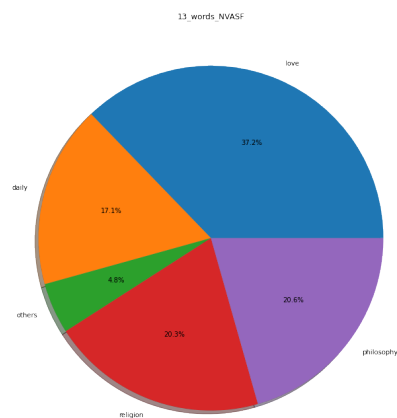
شکل ۵۰: ترکیبات وصفی

۵.



شکل ۵۱: کلمات

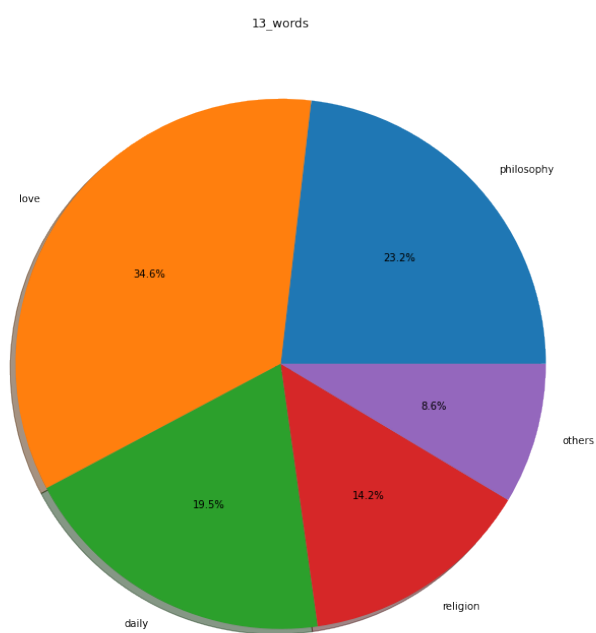
شکل ۵۲: سهم هر مفهوم از کلمات و ترکیبات وصفی و اضافی قرن ۱۲ ام



شکل ۵۳: ترکیبات اضافی

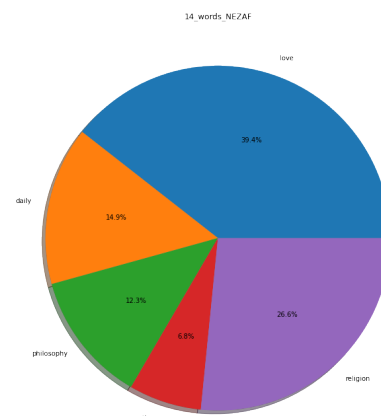
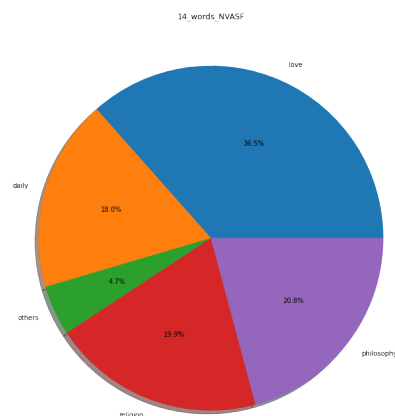
شکل ۵۴: ترکیبات وصفی

۵.



شکل ۵۵: کلمات

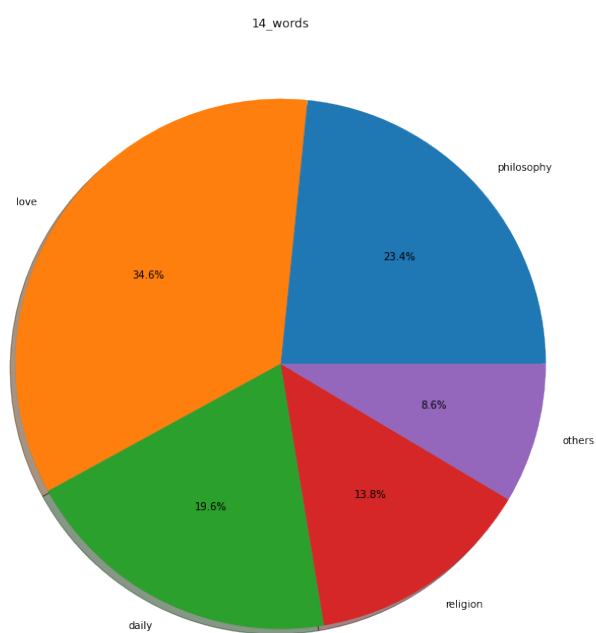
شکل ۵۶: سهم هر مفهوم از کلمات و ترکیبات وصفی و اضافی قرن ۱۳ ام



شکل ۵۷: ترکیبات اضافی

شکل ۵۸: ترکیبات وصفی

۵.



شکل ۵۹: کلمات

شکل ۶۰: سهم هر مفهوم از کلمات و ترکیبات وصفی و اضافی قرن ۱۴ ام

۵ بررسی ۶ شاعر پرمخاطب سایت گنجور و مقایسه آنها با یکدیگر با استفاده از دیتاهای بدست آمده از قرن های ۳ تا ۱۴

در این قسمت شش شاعر و نویسنده پرمخاطب سایت گنجور را بررسی میکنیم.
برای این بخش باید به نوتبوک HW2_main.ipynb مراجعه کنید. همچنین فایل های مربوط به دسته بندی مفاهیم در پوشه هستند.

```
type_dataset/unique_most_NEZAF.csv  
type_dataset/unique_most_NVASF.csv  
type_dataset/unique_most_words.csv
```

برای شش شاعر نام برده شده مراحل ۱ تا ۶ بخش ۱.۳ را تکرار می کنیم.
۲۰۰ کلمه پر تکرار هر شاعر و ۵۰ ترکیب پر تکرار اضافی و ۵۰ ترکیب پر تکرار وصفی را بررسی می کنیم.

دقت کنید که از آنجا که دیتاست برای کلاس بندی مفاهیم بر روی تمام قرون انجام گرفته است بنابراین ممکن است بخشی از ترکیبات بیشتر از حد معمول مربوط به کلاس others باشند. بنابراین برای هر شاعر ۲ حالت را بررسی میکنیم. حالتی که کلاس others را دخیل کرده ایم و حالتی که تنها ۴ کلاس love ، religion ، philosophy و daily را دخیل کرده ایم و کلاس others را کنار گذاشته ایم.

۱.۵ ترکیبات و کلمات پر تکرار

	hafez	saadi	moulavi	ferdousi	khayyam	nezami
0	(یاد صبا, 18)	(غم دل, 19)	(شمس تبریزی, 163)	(شاه جهان, 121)	(دفع مضرتش, 12)	(سر تخت, 26)
1	(یاد باد, 18)	(حدیث عشق, 17)	(شمس تبریز, 146)	(از ان پس, 105)	(زبان حال, 6)	(آب حیوان, 26)
2	(سر زلف, 16)	(خاک پای, 16)	(شمس دین, 130)	(نزدیک شاه, 100)	(چرخ فلک, 5)	(شاه جهان, 22)
3	(خون دل, 14)	(یاد صبا, 14)	(آب حیات, 112)	(جهان آفرین, 90)	(سر سبزه, 4)	(آب حیات, 20)
4	(دل حافظ, 12)	(سر کوی, 14)	(جان جان, 106)	(شاه ایران, 84)	(عبدالله طاهر, 4)	(سر خاک, 19)
5	(پیر مغان, 12)	(دل ریش, 14)	(شمس الدین, 88)	(ایران سپاه, 81)	(رفتن دی, 4)	(زمین بوس, 16)
6	(سر کوی, 10)	(دوست دست, 14)	(شمس الحق, 86)	(دل شاه, 79)	(زیر زمین, 4)	(خاره سنگ, 16)
7	(کام دل, 9)	(سر زلف, 13)	(آب حیوان, 62)	(تخت عاج, 77)	(یزبان پهلوی, 4)	(سهی سرو, 15)
8	(خون جگر, 9)	(دل سعدی, 13)	(تبریز شمس, 50)	(گردش روزگار, 72)	(سال پادشاهی, 4)	(چاره ساز, 14)

شکل ۶۱: ترکیبات اضافی

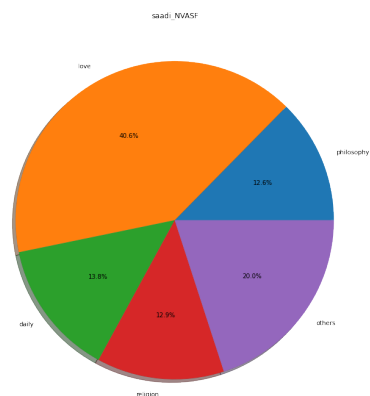
	hafez	saadi	moulavi	ferdousi	khayyam	nezami
0	(چشم بد, 10)	(سرو روان, 19)	(چشم بد, 63)	(شب تیره, 173)	(ملوک عجم, 11)	(چشم بد, 44)
1	(دل خوش, 7)	(یار عزیز, 13)	(آب روان, 34)	(دل پر, 146)	(ایزد تعالی, 6)	(چرخ بلند, 20)
2	(ماه نو, 6)	(خدای عزوجل, 11)	(شیر نر, 29)	(گرز گران, 76)	(روی نیکو, 6)	(گل سرخ, 17)
3	(شاه شجاع, 6)	(خدای عز, 11)	(حق تعالی, 26)	(یزدان پاک, 60)	(دیدار نیکو, 4)	(تنی چند, 17)
4	(تازه نو, 6)	(سرو بلند, 11)	(جهان پر, 23)	(یاد سرد, 58)	(حالی خوش, 4)	(سپهر بلند, 15)
5	(آب جوی, 5)	(روز روشن, 9)	(رطل گران, 21)	(چرخ بلند, 57)	(سبزه نشین, 3)	(چرخ کیود, 15)
6	(مرغ زیرک, 5)	(عمر عزیز, 9)	(لاله زار, 19)	(وزان جایگاه, 52)	(خدای تعالی, 3)	(ماه نو, 15)
7	(عمر دراز, 4)	(یار مهربان, 9)	(رو ترش, 19)	(خورشید تابان, 50)	(خط نیکو, 3)	(چشم روشن, 14)
8	(وادی ایمن, 4)	(جواب تلخ, 9)	(خوی بد, 18)	(دلی پر, 47)	(ایزد تبارک, 3)	(کیک دری, 14)

شکل ۶۲: ترکیبات وصفی

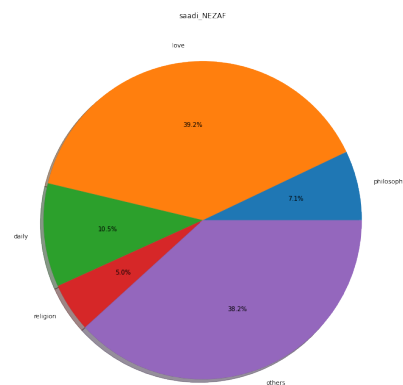
	hafez	saadi	moulavi	ferdousi	khayyam	nezami
0	(حافظ, 509)	(دست, 1312)	(جان, 7718)	(شاه, 4617)	(خاک, 69)	(شاه, 1197)
1	(جان, 306)	(دوست, 884)	(عشق, 3719)	(بد, 2220)	(جهان, 66)	(دست, 1136)
2	(باد, 284)	(جان, 686)	(آب, 2632)	(سپاه, 2124)	(دست, 65)	(کار, 1042)
3	(چشم, 258)	(سعدی, 669)	(چشم, 2347)	(جهان, 2097)	(خوش, 58)	(جهان, 1023)
4	(عشق, 254)	(چشم, 663)	(دست, 2213)	(لشکر, 1805)	(ملک, 53)	(آب, 942)
5	(دست, 242)	(عشق, 591)	(حق, 2124)	(سخن, 1782)	(نیکو, 50)	(سخن, 771)
6	(خوش, 237)	(شب, 500)	(صد, 2096)	(کار, 1594)	(مرد, 50)	(جان, 771)
7	(غم, 218)	(سخن, 466)	(جهان, 2062)	(ایران, 1577)	(ماه, 50)	(خاک, 756)
8	(یار, 210)	(دشمن, 441)	(خوش, 1936)	(مرد, 1519)	(عمر, 49)	(یاد, 732)
9	(زلف, 208)	(غم, 433)	(شب, 1704)	(تخت, 1480)	(زمین, 49)	(شب, 702)
10	(جام, 193)	(ملک, 431)	(مست, 1704)	(جنگ, 1373)	(شراب, 45)	(شه, 626)
11	(آب, 188)	(یار, 401)	(دم, 1553)	(دست, 1353)	(پادشاه, 45)	(ماه, 601)
12	(جهان, 166)	(خلق, 390)	(عقل, 1547)	(زمین, 1265)	(آب, 42)	(گنج, 589)

شکل ۶۳: کلمات

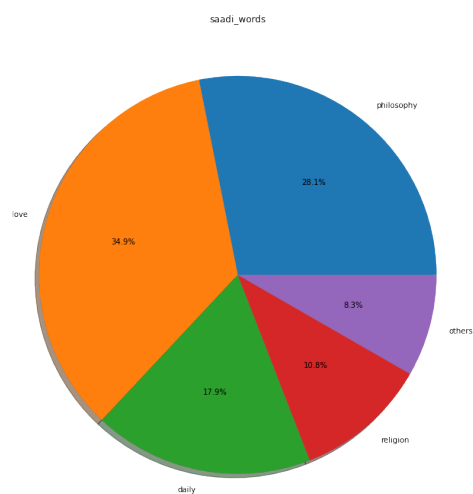
شکل ۶۴: ۶ شاعران پر مخاطب در یک نگاه



شکل ۶۶: ترکیبات وصفی

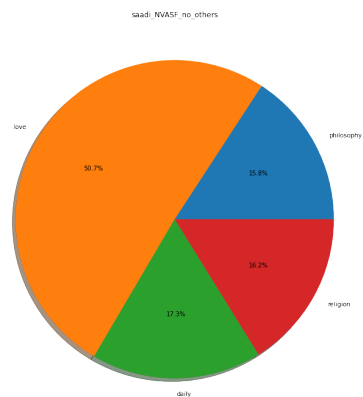


شکل ۶۵: ترکیبات اضافی

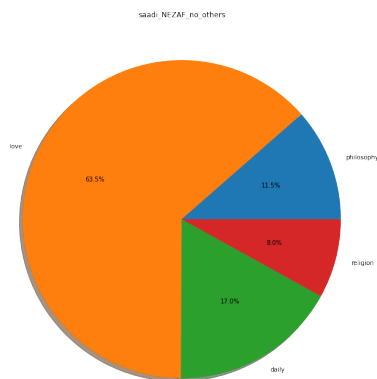


شکل ۶۷: کلمات

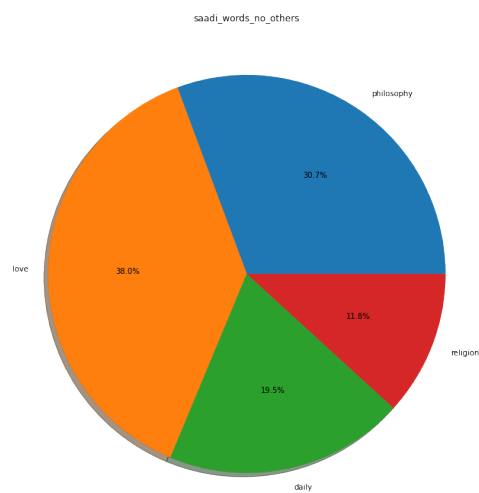
شکل ۶۸: سهم هر مفهوم از کلمات و ترکیبات وصفی و اضافی در آثار سعدی



شکل ۷۰: ترکیبات وصفی



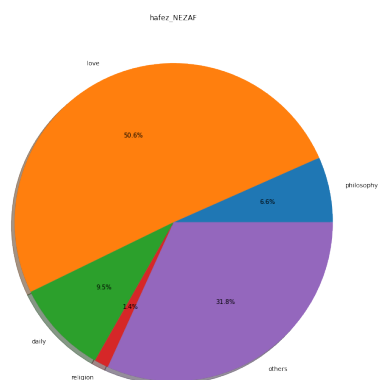
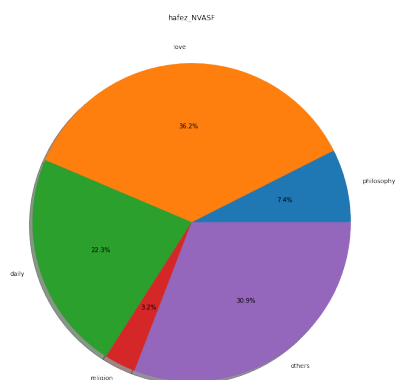
شکل ۶۹: ترکیبات اضافی



شکل ۷۱: کلمات

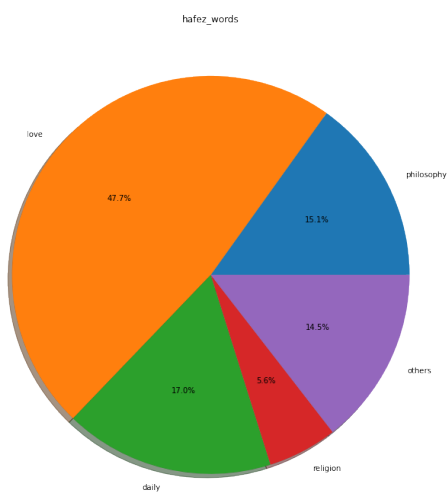
شکل ۷۲: سهم هر مفهوم از کلمات و ترکیبات وصفی و اضافی در آثار سعدی بدون دخیل دادن others

پرداختن سعدی به آثار فلسفی اخلاقی بیشتر بیشتر از میانگین آثار ادبیات فارسی است که در بخش قرن به آن پرداخته شد. اما بقیه مفاهیم اختلاف چندانی با میانگین آثار ادبیات فارسی ندارند.



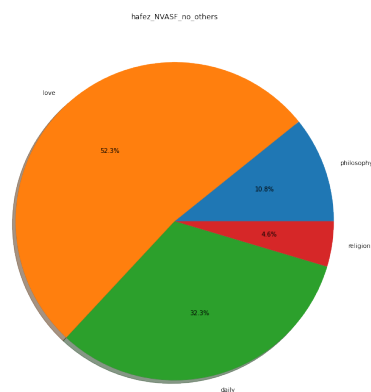
شکل ۷۴: ترکیبات وصفی

شکل ۷۳: ترکیبات اضافی

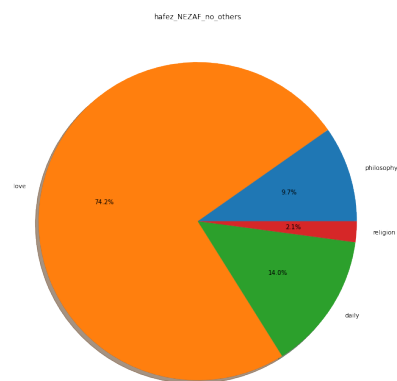


شکل ۷۵: کلمات

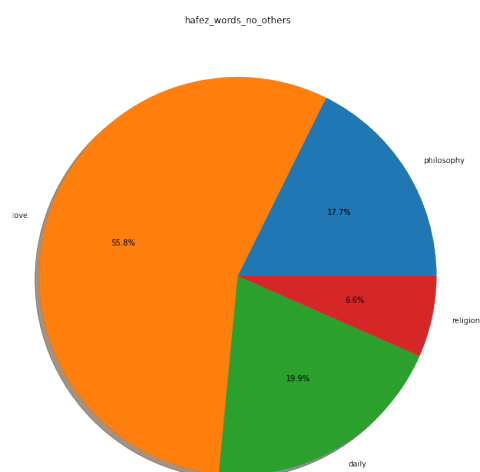
شکل ۷۶: سهم هر مفهوم از کلمات و ترکیبات وصفی و اضافی در آثار حافظ



شکل ۷۸: ترکیبات وصفی



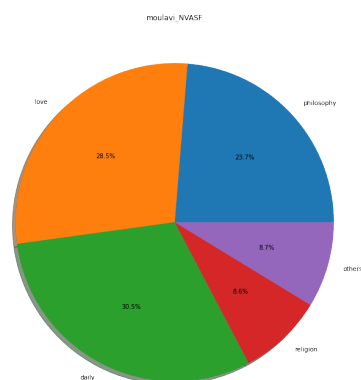
شکل ۷۷: ترکیبات اضافی



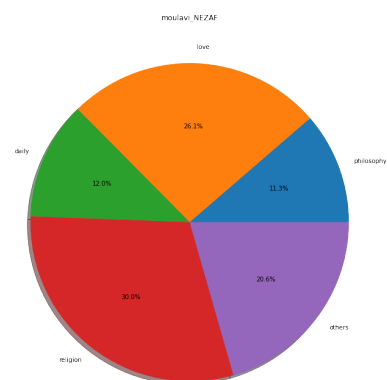
شکل ۷۹: کلمات

شکل ۸۰: سهم هر مفهوم از کلمات و ترکیبات وصفی و اضافی در آثار حافظ بدون دخیل دادن others

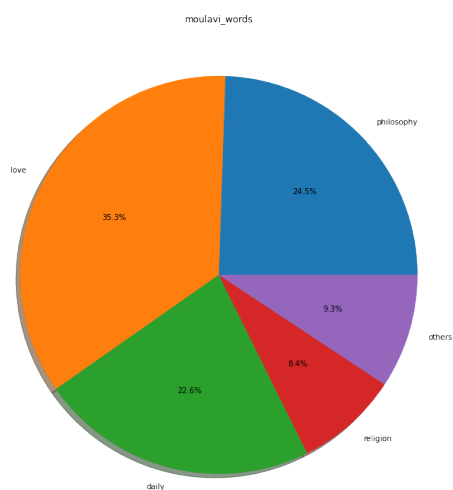
به طور میانگین ۳۴ درصد لغات در ادبیات فارسی به مفاهیم عاشقانه و عارفانه مربوط می شوند اما حافظ نزدیک به ۶۰ درصد از کلمات پرتکرار اثرش به مفاهیم عاشقانه و عارفانه میپردازند. درصد عظیمی از ترکیبات وصفی و اضافی در آثار حافظ به مفاهیم عاشقانه پرداخته اند.



شکل ۸۲: ترکیبات وصفی

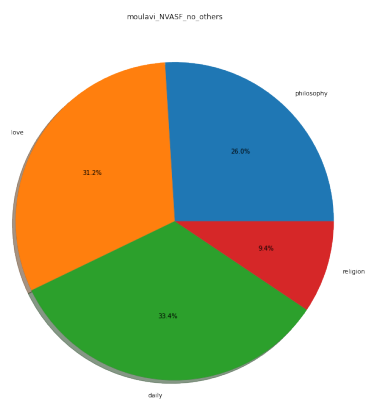


شکل ۸۱: ترکیبات اضافی

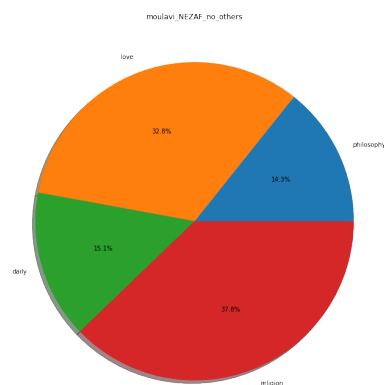


شکل ۸۳: کلمات

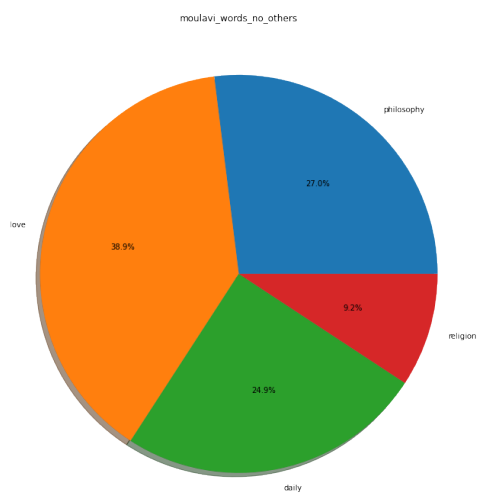
شکل ۸۴: سهم هر مفهوم از کلمات و ترکیبات وصفی و اضافی در آثار مولوی



شکل ۸۶: ترکیبات وصفی



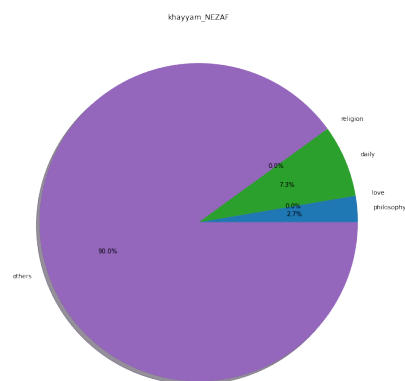
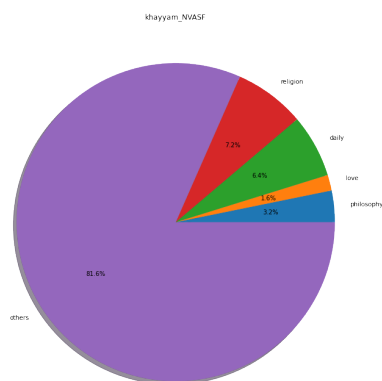
شکل ۸۵: ترکیبات اضافی



شکل ۸۷: کلمات

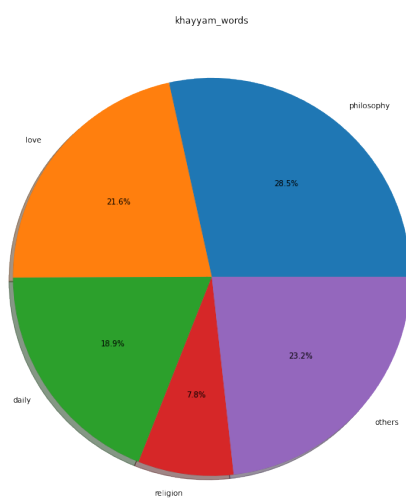
شکل ۸۸: سهم هر مفهوم از کلمات و ترکیبات وصفی و اضافی در آثار مولوی بدون دخیل دادن others

مولوی نیز بیشتر به عشق و فلسفه پرداخته است اما نکته قابل توجه این است که ترکیبات اضافی و وصفی به نسبت بقیه شعرا در ادبیات فارسی بخش بزرگتری از خود را به مفاهیم دینی پرداخته اند همچنین مولوی بیشتر از سایر شعرا بر مخاطب از کلمات مذهبی استفاده کرده است.



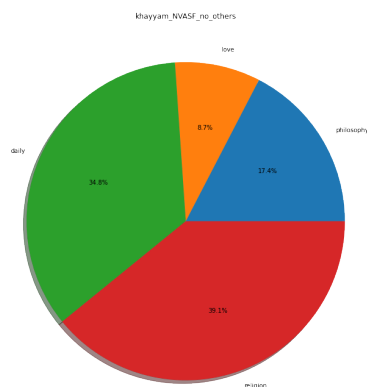
شکل ۹۰: ترکیبات وصفی

شکل ۸۹: ترکیبات اضافی

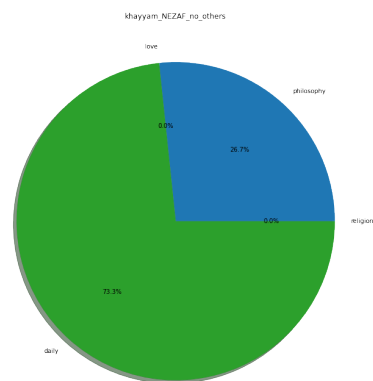


شکل ۹۱: کلمات

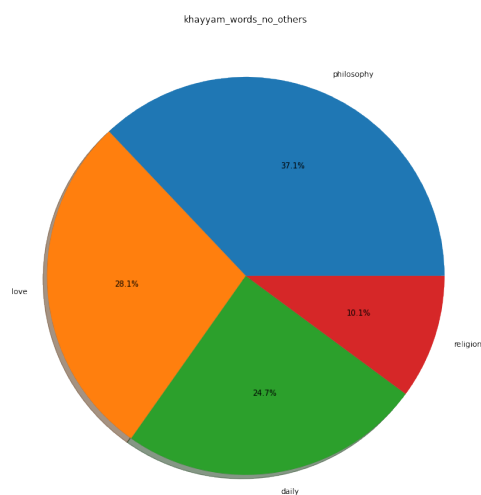
شکل ۹۲: سهم هر مفهوم از کلمات و ترکیبات وصفی و اضافی در آثار خیام



شکل ۹۴: ترکیبات وصفی



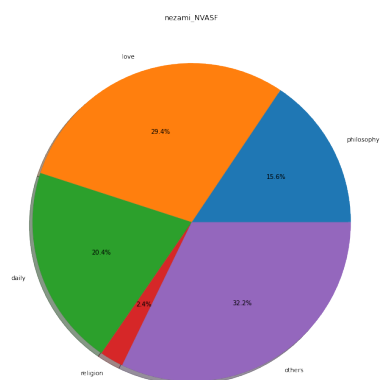
شکل ۹۳: ترکیبات اضافی



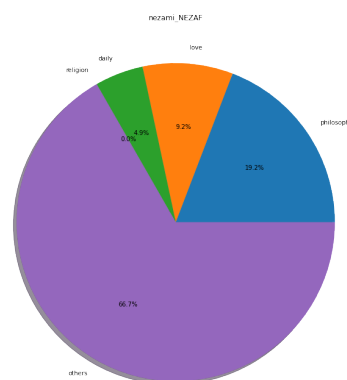
شکل ۹۵: کلمات

شکل ۹۶: سهم هر مفهوم از کلمات و ترکیبات وصفی و اضافی در آثار خیام بدون دخیل دادن others

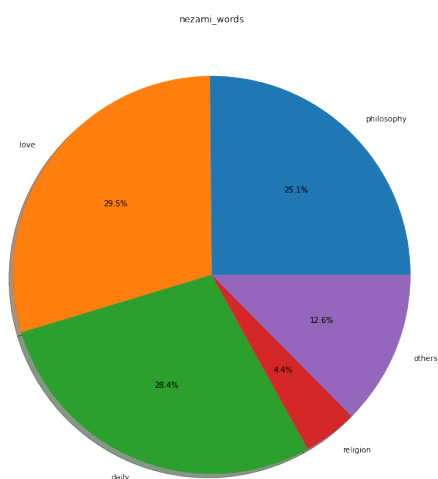
تکته قابل توجه در آثار خیام پرداختن زیاد وی به مفاهیم روزمره و عادی مانند شادی و شراب (از نوع زمینی) روزمرگی در آثار خیام نقش قابل توجه ای دارند به عنوان مثال بخش عظیمی از کلمات و ترکیبات اضافی در آثار خیام به مفاهیم روزمره و زندگی زمینی و عادی پرداخته اند.



شکل ۹۸: ترکیبات وصفی

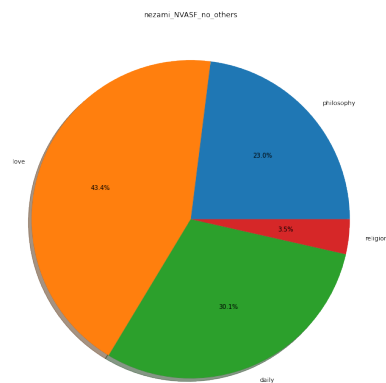


شکل ۹۷: ترکیبات اضافی

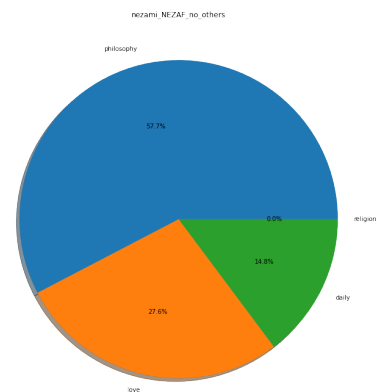


شکل ۹۹: کلمات

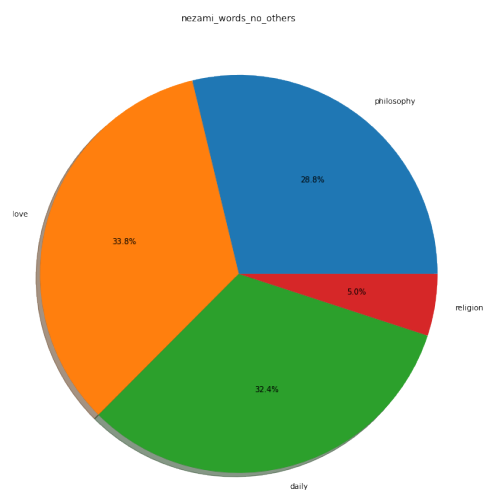
شکل ۱۰۰: سهم هر مفهوم از کلمات و ترکیبات وصفی و اضافی در آثار نظامی



شکل ۱۰۲: ترکیبات وصفی



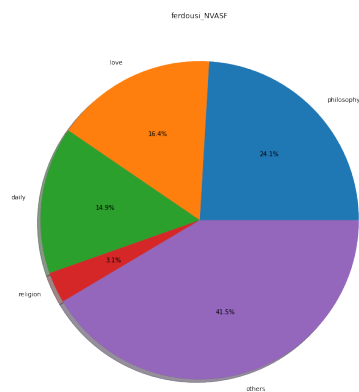
شکل ۱۰۱: ترکیبات اضافی



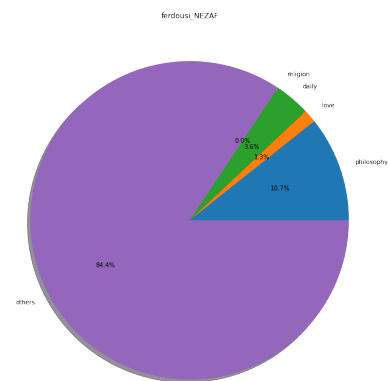
شکل ۱۰۳: کلمات

شکل ۱۰۴: سهم هر مفهوم از کلمات و ترکیبات وصفی و اضافی در آثار نظامی بدون دخیل دادن others

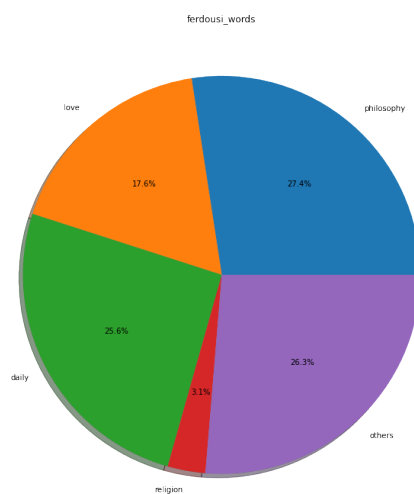
سراینده لیلی و معجون و خسرو و شیرین به ندرت از ترکیبات با بار مذهبی استفاده نموده است همچنین نظامی به مساءل عاشقانه و فلسفی پرداخته است.. پرداختن به مساءل روزمرگی نیز بخش قابل توجه ای از کلمات و ترکیبات استفاده شده توسط این شاعر بزرگوار است.



شکل ۱۰۶: ترکیبات وصفی

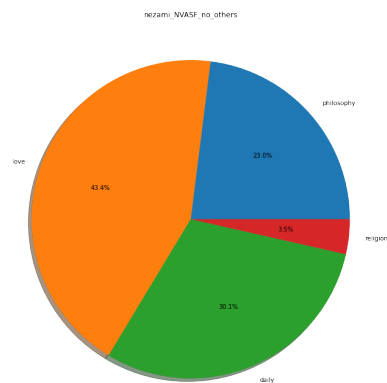


شکل ۱۰۵: ترکیبات اضافی

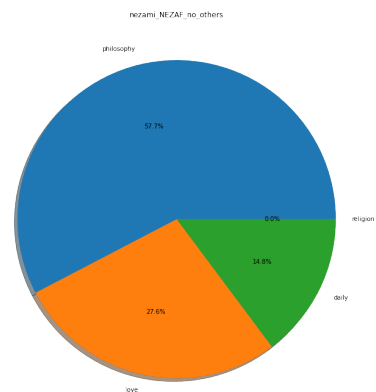


شکل ۱۰۷: کلمات

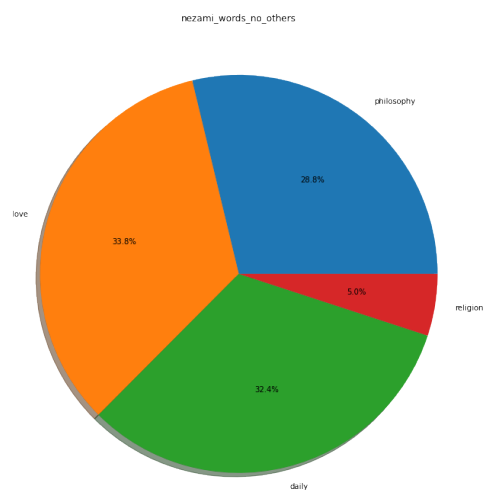
شکل ۱۰۸: سهم هر مفهوم از کلمات و ترکیبات وصفی و اضافی در آثار فردوسی



شکل ۱۱۰: ترکیبات وصفی



شکل ۱۰۹: ترکیبات اضافی



شکل ۱۱۱: کلمات

شکل ۱۱۲: سهم هر مفهوم از کلمات و ترکیبات وصفی و اضافی در آثار فردوسی بدون دخیل دادن others

دو نکته درباره فردوسی به چشم میخورد اول آنکه درصد ترکیبات وصفی و اضافی فلسفی به نسبت زیاد است زیرا همانطور که مطرح شد مفاهیم سیاسی و اخلاقی را نیز در دسته فلسفی قرار داده ایم. همچنین فردوسی به ندرت به مفاهیم دینی پرداخته است و در کنار آن فردوسی به مسائل عاشقانه نیز اندکی پرداخته است.