

# تحلیل و بررسی مقالات علمی نوشته شده در ایران

پرند علیزاده ۹۴۱۰۰۰۲۴

## مقدمه

سالانه در دانشگاه ها و موسسات علمی جهام مقالات علمی زیادی در موضوعات مختلف تالیف و منتشر میشوند و طبق تخمین *nature* این مقدار هر ۹ سال دو برابر می شود. [1]

## اهداف

در این پروژه هدف بررسی این مقالات، پاسخ به کنجکاوی های مختلف راجع به آنها و به دست آوردن ویژگی های آنها است. تمرکز اصلی این پروژه روی مقالان ایران است و تحلیل و بررسی آنها با بدست آوردن نقاط ضعف و قوت آنها میتواند باعث افزایش کیفیت و برطرف کردن مشکلات و ایرادات مقالات علمی در ایران شود.

## کنجکاوی ها

چند کنجکاوی اولیه در حالت کلی بیان شده است:

تحلیل اولیه ویژگی های مقالات (مانند ویژگی های متنی)

تحلیل زمانی مقالات

تحلیل مکانی مقالات

تحلیل شبکه دانشگاه ها و موسسه های مختلف

تحلیل شبکه نویسندگان مقالات علمی

تحلیل موضوعات مقالات در طول زمان

تحلیل اثرات موضوعات مقاله بر ویژگی های متنی آن

تحلیل اثرات نویسندگان مقاله و دانشگاه ها بر ویژگی های متنی آن

(اضافی و اختیاری : در نهایت میتوان موارد بالا را در مقالات خارجی نیز بررسی نمود و نتایج را باهم مقایسه کرد. این قسمت خارج از قسمت اصلی پروژه است و برای پاسخ به کنجکاوی های بیشتر، در صورت داشتن زمان کافی بررسی میگردد.)

## منابع پروژه

داده های مربوط به مقالات علمی ایران قبلا استخراج شده و موجود هستند.

(داده های مربوط به قسمت اضافی آخر (مقالات خارجی) در برخی موضوعات و سالها در سایت ها موجود هستند و همچنین در صورت نیاز میتوان داده ها را از سایت [arxiv.org](http://arxiv.org) جمع آوری کرد.)

## ابزارهای انجام پروژه

ابزار اصلی تحلیل داده های پروژه زبان R و package های موجود در آن است.

(در صورت انجام قسمت اختیاری، میتوان برای پردازش داده های حجیم از [2] Apache Spark و زبان scala یا از ابزار تعبیه شده در R برای Spark، rspark استفاده کرد.)

## References

[1] <http://blogs.nature.com/news/2014/05/global-scientific-output-doubles-every-nine-years.html>

[2] <https://spark.apache.org>