



دانشگاه صنعتی شریف
دانشکده علوم ریاضی

کارشناسی ارشد
علوم کامپیوتر

قالب استاندارد برای نگارش پایان نامه ها

نگارش
نام دانشجو

استاد راهنما
استاد راهنمای پروژه

شهریور ۱۴۰۴

تصویب نامه

دانشگاه صنعتی شریف
دانشکده علوم ریاضی

پایان نامه ی کارشناسی ارشد

این پایان نامه به عنوان تحقق بخشی از شرایط دریافت درجه ی کارشناسی ارشد است.

عنوان: قالب استاندارد برای نگارش پایان نامه ها
نگارش: نام دانشجو

کمیته ممتحنین

استاد راهنما: استاد راهنمای پروژه امضا:

استاد مشاور: استاد مشاور امضا:

استاد مدعو: استاد ممتحن امضا:

تاریخ:

اظهارنامه

اصالت متن و محتوای پایان نامه

عنوان پایان نامه: قالب استاندارد برای نگارش پایان نامه ها

استاد راهنما: استاد راهنمای پروژه

استاد مشاور: استاد مشاور

اینجانب نام دانشجو اظهار می دارم که

۱. متن و نتایج علمی ارایه شده در این پایان نامه اصیل بوده و زیر نظر استادان نام برده شده در بالا تهیه شده است.

۲. متن پایان نامه به این صورت در هیچ جای دیگری منتشر نشده است.

۳. متن و نتایج مندرج در این پایان نامه، حاصل تحقیقات این جانب به عنوان دانشجوی کارشناسی ارشد دانشگاه صنعتی شریف است.

۴. کلیه مطالبی که از منابع دیگر در این پایان نامه مورد استفاده قرار گرفته، با ذکر مرجع مشخص شده است.

نگارنده: نام دانشجو

تاریخ:

امضا:

نتایج تحقیقات مندرج در این پایان نامه و دستاوردهای مادی و معنوی ناشی از آن (شامل فرمول ها، توابع کتابخانه ای، نرم افزارها، سخت افزارها و مواردی که قابلیت ثبت اختراع دارد) متعلق به دانشگاه صنعتی شریف است. هیچ شخصیت حقیقی یا حقوقی بدون کسب اجازه از دانشگاه صنعتی شریف حق فروش و ادعای مالکیت مادی یا معنوی بر آن یا ثبت اختراع از آن را ندارد. همچنین، کلیه حقوق مربوط به چاپ، تکثیر، نسخه برداری، ترجمه، اقتباس و نظایر آن در محیط های مختلف اعم از الکترونیکی، مجازی یا فیزیکی برای دانشگاه صنعتی شریف محفوظ است. نقل مطلب با ذکر ماخذ بلامانع است.

نگارنده: نام دانشجو

تاریخ:

امضا:

استاد راهنما: استاد راهنمای پروژه

تاریخ:

امضا:

سپاس‌گزاری

از استاد بزرگوارم که با کمک‌ها و راهنمایی‌های بی‌دریغشان مرا در به سرانجام رساندن این پایان‌نامه یاری داده‌اند
تشکر و قدردانی می‌کنم. همچنین از همکاران عزیزی که با راهنمایی‌های خود در بهبود نگارش این نوشتار سهیم بوده‌اند،
صمیمانه سپاس‌گزارم.

چکیده

نگارش پایان‌نامه علاوه بر بخش پژوهش و آماده‌سازی محتوا، مستلزم رعایت نکات فنی و نگارشی دقیقی است که در تهیه‌ی یک پایان‌نامه‌ی موفق بسیار کلیدی و موثر است. از آن جایی که بسیاری از نکات فنی مانند قالب کلی صفحات، شکل و اندازه‌ی قلم، صفحات عنوان و غیره در تهیه‌ی پایان‌نامه‌ها یکسان است، با استفاده از نرم‌افزار حروف‌چینی زی‌لاتک^۱ و افزونه‌ی زی‌پرشین^۲ یک قالب استاندارد برای تهیه‌ی پایان‌نامه‌ها ارایه گردیده است. زی‌پرشین توسط دکتر وفا خلیقی توسعه داده شده است و در حال حاضر بهترین افزونه‌ی پارسی برای تهیه‌ی متون در لاتک^۳ به شمار می‌رود. این قالب می‌تواند برای تهیه‌ی پایان‌نامه‌های کارشناسی و کارشناسی ارشد و نیز رساله‌های دکتری مورد استفاده قرار گیرد. این نوشتار به طور مختصر نحوه‌ی استفاده از این قالب را نشان می‌دهد. قالب اولیه توسط دکتر ضرابی‌زاده و دانشجویان دانشکده‌ی مهندسی کامپیوتر شریف تهیه شد که از **مخزن کد** آن در GitHub قابل دریافت است. این نسخه توسط حسین رهنما ویرایش و شخصی‌سازی شده است. تغییرات عمده در این ویرایش حذف دستورات پارسی در کد لاتک، تغییر ساختار فایل‌ها و پوشه‌ها، بهبود کدها و بهبود ظاهر قالب است.

کلیدواژه‌ها. پایان‌نامه، حروف‌چینی، قالب استاندارد، لاتک، زی‌تک، زی‌پرشین.

^۱ \LaTeX

^۲ \LaTeX Persian

^۳ \LaTeX

فهرست مطالب

۱	۱ مقدمه
۱	۱.۱ تعریف مساله
۱	۲.۱ اهمیت موضوع
۱	۳.۱ ادبیات موضوع
۲	۴.۱ اهداف پژوهش
۲	۵.۱ ساختار پایان نامه
۳	۲ مفاهیم اولیه
۳	۱.۲ نحوه نگارش
۳	۱.۱.۲ ساختار فایل ها و پوشه ها
۳	۲.۱.۲ عبارات ریاضی
۴	۳.۱.۲ نمادهای ریاضی پرکاربرد
۴	۴.۱.۲ لیست ها
۴	۵.۱.۲ درج شکل
۵	۶.۱.۲ درج جدول
۵	۷.۱.۲ درج الگوریتم
۵	۸.۱.۲ محیط های ویژه
۶	۲.۲ برخی نکات نگارشی
۶	۱.۲.۲ فاصله گذاری
۷	۲.۲.۲ شکل حروف
۷	۳.۲.۲ جدانویسی
۸	۳ کارهای پیشین
۸	۱.۳ مساله های خوشه بندی
۹	۲.۳ خوشه بندی k -مرکز
۱۱	۳.۳ مدل جویبار داده
۱۲	۴.۳ تقریب پذیری
۱۳	۴ نتایج جدید
۱۴	۵ نتیجه گیری
۱۵	آ مطالب تکمیلی

مراجع

واژه‌نامه

۱۶

۱۷

فهرست جدول‌ها

۵	۱.۲ عملگرهای مقایسه‌ای
۱۲	۲.۳ نمونه‌هایی از کران پایین تقریب‌پذیری مسائل خوشه‌بندی

فهرست شکل‌ها

۵	۱.۲ نمونه شکل ایجاد شده توسط نرم‌افزار LaTeX Draw
۵	۲.۲ نمونه شکل ایجاد شده توسط نرم‌افزار lpe
۹	۳.۳ نمونه‌ای از مساله‌ی ۲-مرکز
۱۰	۴.۳ نمونه‌ای از مساله‌ی ۲-مرکز با داده‌های پرت

فهرست الگوریتم‌ها

۱.۲ پوشش راسی حریصانه	۶
-----------------------	---

فصل ۱

مقدمه

نخستین فصل یک پایان‌نامه به معرفی مسئله، بیان اهمیت موضوع، ادبیات موضوع، اهداف پژوهش و معرفی ساختار پایان‌نامه می‌پردازد. در این فصل نمونه‌ی مختصری از مقدمه آورده شده است.

۱.۱ تعریف مساله

نگارش یک پایان‌نامه علاوه بر بخش‌های پژوهش و آماده‌سازی محتوا، مستلزم رعایت نکات دقیق فنی و نگارشی است که در تهیه‌ی یک پایان‌نامه‌ی موفق بسیار کلیدی و موثر است. از آن جایی که بسیاری از نکات فنی مانند قالب کلی صفحات، شکل و اندازه‌ی قلم، صفحات عنوان و غیره در تهیه‌ی پایان‌نامه‌ها یکسان است، می‌توان با ارائه‌ی یک قالب حروف‌چینی استاندارد نگارش پایان‌نامه‌ها را تا حد بسیار زیادی بهبود بخشید.

۲.۱ اهمیت موضوع

- وجود قالب استاندارد برای نگارش پایان‌نامه از جهات مختلف حائز اهمیت است، از جمله:
- ایجاد یک‌نواختی در قالب کلی صفحات و شکل و اندازه‌ی قلم‌ها
 - تسهیل نگارش پایان‌نامه با در اختیار گذاشتن یک قالب اولیه
 - تولید خودکار صفحات دارای بخش‌های تکراری نظیر صفحات ابتدایی و انتهایی پایان‌نامه
 - پیش‌گیری از برخی خطاهای مرسوم در نگارش پایان‌نامه

۳.۱ ادبیات موضوع

اکثر دانشگاه‌ها قالب استاندارد برای تهیه‌ی پایان‌نامه‌ها در اختیار دانشجویان خود قرار می‌دهند. این قالب‌ها عموماً مبتنی بر نرم‌افزارهای متداول حروف‌چینی نظیر لاتک و مایکروسافت ورد^۱ هستند. لاتک یک نرم‌افزار متن‌باز قوی برای حروف‌چینی متون علمی است [۱، ۲]. در این نوشتار از نرم‌افزار حروف‌چینی زی‌تک و افزونه‌ی زی‌پرشین استفاده شده است.

^۱Microsoft Word

۴.۱ اهداف پژوهش

کتابخانه‌ی مرکزی دانشگاه صنعتی شریف دستورالعمل جامعی در خصوص نحوه‌ی تهیه‌ی پایان‌نامه‌های کارشناسی ارشد و رساله‌های دکتری ارائه کرده است. در این نوشتار سعی شده است قالب استاندارد برای تهیه‌ی پایان‌نامه‌ها مبتنی بر نرم‌افزار لاتک و براساس دستورالعمل مذکور ارائه شده و نحوه‌ی استفاده از قالب به طور مختصر توضیح داده شود. این قالب می‌تواند برای تهیه‌ی پایان‌نامه‌های کارشناسی، کارشناسی ارشد و همچنین رساله‌های دکتری مورد استفاده قرار گیرد.

۵.۱ ساختار پایان‌نامه

این پایان‌نامه در پنج فصل به شرح زیر ارائه می‌شود. مفاهیم اولیه مورد استفاده در این پایان‌نامه در فصل دوم به اختصار اشاره شده است. فصل سوم به مطالعه و بررسی کارهای پیشین مرتبط با موضوع این پایان‌نامه می‌پردازد. در فصل چهارم، نتایج جدیدی که در این پایان‌نامه به دست آمده است، ارائه می‌شود. فصل پنجم به جمع‌بندی کارهای انجام شده در این پژوهش و ارائه‌ی پیشنهادهایی برای انجام کارهای آتی خواهد پرداخت.

فصل ۲

مفاهیم اولیه

دومین فصل پایان نامه به طور معمول به معرفی مفاهیمی می پردازد که در پایان نامه مورد استفاده قرار می گیرند. در این فصل به عنوان یک نمونه، نکات کلی در خصوص نحوه نگارش پایان نامه و نیز برخی نکات نگارشی به اختصار توضیح داده می شوند.

۱.۲ نحوه نگارش

۱.۱.۲ ساختار فایل ها و پوشه ها

فایل اصلی پایان نامه در این قالب thesis.tex نام دارد. به ازای هر فصل از پایان نامه، یک پوشه در شاخه ی body/chapters-x ایجاد شده است و نام آن در thesis.tex (در قسمت فصل ها) درج گردیده است. برای مشاهده ی خروجی، فایل thesis.tex را با زی لاتک و بیب تک^۱ کامپایل کنید. مشخصات اصلی پایان نامه به زبان پارسی را می توانید در پرونده ی body/front/info.tex ویرایش کنید. تمامی محتویاتی که در پایان نامه ی خود باید درج کنید در پوشه ی body قرار گرفته است. پوشه ی styles شامل پکیج های مورد استفاده و تنظیمات این قالب است که در صورت نیاز می توانید آن را تغییر دهید. تمامی عکس ها در پوشه ی figs قرار دارند. هم چنین پوشه ی fonts دربرگیرنده ی فونت های مورد استفاده در این قالب است.

۲.۱.۲ عبارات ریاضی

برای درج عبارات ریاضی در داخل متن از $\$... \$$ و برای درج عبارات ریاضی در یک خط مجزا از $\$ \$... \$ \$$ یا محیط equation استفاده کنید. برای مثال عبارت $2x + 3y$ در داخل متن و عبارت زیر

$$\sum_{k=0}^n \binom{n}{k} = 2^n \quad (۱.۲)$$

در یک خط مجزا درج شده است. دقت کنید که تمامی عبارات ریاضی، از جمله متغیرهای تک حرفی مانند x و y باید در محیط ریاضی یعنی محصور بین دو علامت $\$$ باشند.

¹BibTeX

۳.۱.۲ نمادهای ریاضی پرکاربرد

برخی نمادهای ریاضی پرکاربرد در زیر فهرست شده‌اند. برای مشاهده‌ی دستور معادل پرونده‌ی منبع را ببینید.

- مجموعه‌های اعداد: $\mathbb{N}, \mathbb{Z}, \mathbb{Z}^+, \mathbb{Q}, \mathbb{R}, \mathbb{C}$
- مجموعه: $\{1, 2, 3\}$
- دنباله: $\langle 1, 2, 3 \rangle$
- سقف و کف: $\lceil x \rceil, \lfloor x \rfloor$
- اندازه و متمم: $|A|, \bar{A}$
- همنهشتی: $a \equiv 1$ یا $a \equiv 1 \pmod{n}$ (پیمانه‌ی n)
- شمردن (عاد کردن): $3 \mid n, 2 \nmid n$
- ضرب و تقسیم: \times, \cdot, \div
- سه نقطه: $1, 2, \dots, n$
- کسر و ترکیب: $\frac{n}{k}, \binom{n}{k}$
- اجتماع و اشتراک: $A \cup (B \cap C)$
- عملگرهای منطقی: $\neg p \vee (q \wedge r)$
- پیکان‌ها: $\rightarrow, \Rightarrow, \leftarrow, \Leftarrow, \leftrightarrow, \Leftrightarrow$
- عملگرهای مقایسه‌ای: $\neq, \leq, \not\leq, \geq, \not\geq$
- عملگرهای مجموعه‌ای: $\in, \notin, \setminus, \subset, \subseteq, \subsetneq, \supset, \supseteq, \supsetneq$
- جمع و ضرب چندتایی: $\sum_{i=1}^n a_i, \prod_{i=1}^n a_i$
- اجتماع و اشتراک چندتایی: $\bigcup_{i=1}^n A_i, \bigcap_{i=1}^n A_i$
- برخی نمادها: $\infty, \emptyset, \forall, \exists, \Delta, \angle, \ell, \equiv, \therefore$

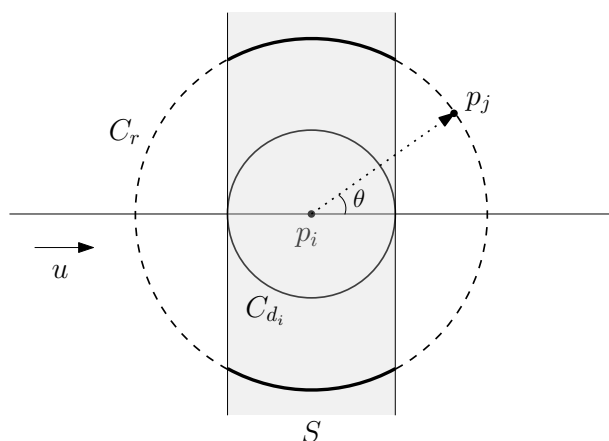
۴.۱.۲ لیست‌ها

برای ایجاد یک لیست می‌توانید از محیط‌های `enumerate` و `itemize` همانند زیر استفاده کنید.

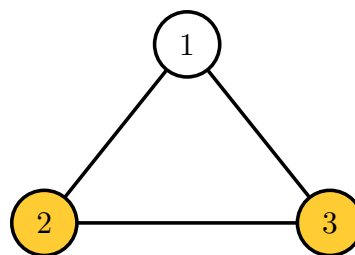
- | | |
|------------|-------------|
| • مورد اول | ۱. مورد اول |
| • مورد دوم | ۲. مورد دوم |
| • مورد سوم | ۳. مورد سوم |

۵.۱.۲ درج شکل

یکی از روش‌های مناسب برای ایجاد شکل استفاده از نرم‌افزار LaTeX Draw و سپس گرفتن خروجی آن به صورت یک فایل `tex`، کامپایل کردن آن به صورت یک فایل `pdf` و قرار دادن آن درون متن با استفاده از محیط `figure` است. همچنین می‌توانید با استفاده از نرم‌افزار `lpe` شکل‌های خود را مستقیماً به صورت `pdf` ایجاد نموده و آن‌ها را به صورت عکس درون متن درج کنید. از بسته‌ی `TikZ` هم می‌توانید برای کشیدن طیف وسیعی از اشکال و به طور خاص گراف‌ها استفاده کنید. قرار دادن دو شکل در کنار هم به گونه‌ای که ترتیب آن‌ها در فهرست شکل‌ها حفظ شود کمی دردسر زاست و باید از روش‌های خاصی برای آن استفاده کرد. برای این منظور از محیط مخصوص `rowfig` استفاده کنید. برای نمونه، شکل ۱.۲ و شکل ۲.۲ با این روش کنار هم قرار گرفته‌اند.



شکل ۲.۲. نمونه شکل ایجادشده توسط نرم افزار lpe



شکل ۱.۲. نمونه شکل ایجاد شده توسط نرم افزار LaTeX Draw

۶.۱.۲ درج جدول

برای درج جدول می‌توانید با استفاده از دستور tabular جدول را ایجاد کرده و سپس با دستور table آن را درون متن درج کنید. برای نمونه جدول ۱.۲ را ببینید.

۷.۱.۲ درج الگوریتم

برای درج الگوریتم می‌توانید از محیط alg استفاده کنید. یک نمونه در الگوریتم ۱.۲ آمده است. همانطور که مشاهده می‌کنید کلمات کلیدی به رنگ آبی تیره و کامنت‌ها به رنگ سبز نمایش داده می‌شوند.

۸.۱.۲ محیط‌های ویژه

برای درج مثال‌ها، قضیه‌ها، لم‌ها و نتیجه‌ها به ترتیب از محیط‌های exmp، thm، lem و cor استفاده کنید. برای درج اثبات قضیه‌ها و لم‌ها از محیط prf استفاده کنید. تعریف‌های داخل متن را با استفاده از دستور به صورت تیره نشان دهید. تعریف‌های پایه‌ای‌تر را درون محیط defn قرار دهید.

تعریف ۱.۲ (اصل لانه‌ی کبوتری). اگر $n + 1$ کبوتر یا بیش‌تر درون n لانه قرار گیرند، آن‌گاه لانه‌ای وجود دارد که شامل حداقل دو کبوتر است.

در هندسه‌ی مسطحه قضیه‌ی زیر را داریم.

گزاره ۱.۲ (فیثاغورس). اگر اعداد حقیقی مثبت $a < b < c$ سه ضلع یک مثلث قائم‌الزاویه باشند، آن‌گاه $c^2 = a^2 + b^2$.

جدول ۱.۲. عملگرهای مقایسه‌ای

عنوان	عملگر
کوچک‌تر	<
بزرگ‌تر	>
مساوی	==
نامساوی	<>

ورودی: گراف $G = (V, E)$

خروجی: یک پوشش راسی از G

۱: تابع SPANNING-VERTICES:

۲: قرار بده $C = \emptyset$

۳: تا وقتی $(E \neq \emptyset)$ انجام بده

۴: اگر $(|E| > 10)$ آن گاه

۵: یک کاری انجام بده

۶: وگرنه اگر $(|E| \leq 5)$ آن گاه

۷: کار دیگری انجام بده

۸: وگرنه

۹: ویه کار دیگه!

۱۰: یال دل خواه $(u, v) \in E$ را انتخاب کن

۱۱: تمام یال های واقع بر u یا v را از E حذف کن

۱۲: * این یک کامنت است.

۱۳: راس های u و v را به C اضافه کن

۱۴: برگردان C

توجه. در گزاره‌ی بالا اگر دو ضلع کوچک‌تر مثلث برابر باشند یعنی $a = b$ ، آن گاه داریم $c = \sqrt{2}a$.

۲.۲ برخی نکات نگارشی

این فصل حاوی برخی نکات ابتدایی ولی بسیار مهم در نگارش متون فارسی است. نکات گردآوری شده در این فصل به هیچ وجه کامل نیست، ولی دربردارنده‌ی حداقل مواردی است که رعایت آن‌ها در نگارش پایان‌نامه ضروری به نظر می‌رسد.

۱.۲.۲ فاصله‌گذاری

۱. علامت‌های سجاوندی مانند نقطه، ویرگول، دونقطه، نقطه‌ویرگول، علامت سؤال و علامت تعجب بدون فاصله از کلمه‌ی پیشین خود نوشته می‌شوند، ولی بعد از آن‌ها باید یک فاصله قرار گیرد. مانند: من، تو، او، چرا؟ وای!
۲. علامت‌های پرانتز، آکولاد، کروشه، نقل قول و نظایر آن‌ها بدون فاصله با عبارات داخل خود نوشته می‌شوند، ولی با عبارات اطراف خود یک فاصله دارند. مانند: (این عبارت) یا {آن عبارت}.
۳. دو کلمه‌ی متوالی در یک جمله همواره با یک فاصله از هم جدا می‌شوند، ولی اجزای یک کلمه‌ی مرکب باید با نیم‌فاصله^۱ از هم جدا شوند. مانند: کتاب درس، محبت آمیز، دویخشی.
۴. اجزای فعل‌های مرکب با فاصله از یک‌دیگر نوشته می‌شوند، مانند: تحریر کردن، به سر آمدن.

^۱ «نیم‌فاصله» فاصله‌ای مجازی است که در عین جدا کردن اجزای یک کلمه‌ی مرکب از یک‌دیگر، آن‌ها را نزدیک به هم نگه می‌دارد. معمولاً برای تولید این نوع فاصله در صفحه‌کلیدهای استاندارد از ترکیب Shift+Space استفاده می‌شود.

۲.۲.۲ شکل حروف

۱. در متون فارسی به جای حروف «ك» و «ي» عربی باید از حروف «ک» و «ی» فارسی استفاده شود. همچنین به جای اعداد عربی مانند ۵ و ۶ باید از اعداد فارسی مانند ۵ و ۶ استفاده نمود. برای این کار، توصیه می‌شود صفحه‌کلید فارسی استاندارد را روی سیستم خود فعال کنید.
۲. عبارات نقل‌قول‌شده یا مؤکد باید درون علامت نقل قول «()» قرار گیرند، نه «”» . مانند: «کشور ایران».
۳. کسره‌ی اضافه‌ی بعد از «ه» غیرملفوظ به صورت «ه‌ی» یا «ه» نوشته می‌شود. مانند: خانه‌ی علی، دنباله‌ی فیبوناچی. اگر «ه» ملفوظ باشد، نیاز به «ی» ندارد. مانند: فرمانده دلیر، پادشه‌ی خوبان.
۴. در این نوشتار ترجیح نویسنده این است که از همزه به هر شکلی استفاده نکند.

۳.۲.۲ جدانویسی

۱. علامت استمرار «می» توسط نیم‌فاصله از جز بعدی فعل جدا می‌شود، مانند: می‌رود، می‌توانیم.
۲. شناسه‌های «ام»، «ای»، «ایم»، «اید» و «اند» توسط نیم‌فاصله و شناسه‌ی «است» توسط فاصله از کلمه‌ی پیش از خود جدا می‌شوند، مانند: گفته‌ام، گفته‌ای، گفته است.
۳. علامت جمع «ها» توسط نیم‌فاصله از کلمه‌ی پیش از خود جدا می‌شود. مانند: این‌ها، کتاب‌ها.
۴. «به» همیشه جدا از کلمه‌ی بعد از خود نوشته می‌شود، مانند: به نام و به آن‌ها، مگر در مواردی که «ب» صفت یا فعل ساخته است، مانند: بسزا، بینم.
۵. «به» همواره با فاصله از کلمه‌ی بعد از خود نوشته می‌شود، مگر در مواردی که «به» جزئی از یک اسم یا صفت مرکب است، مانند: تناظر یک‌به‌یک، سفر به تاریخ.
۶. علامت صفت برتری «تر» و علامت صفت برترین «ترین» توسط نیم‌فاصله از کلمه‌ی پیش از خود جدا می‌شوند، مانند: سنگین‌تر، مهم‌ترین. کلمات «بهتر» و «بهترین» را می‌توان از این قاعده مستثنی نمود.
۷. پیشوندها و پسوندهای جامد، چسبیده به کلمه‌ی پیش یا پس از خود نوشته می‌شوند، مانند: همسر، دانشکده، دانشگاه. در مواردی که خواندن کلمه دچار اشکال می‌شود، می‌توان پسوند یا پیشوند را جدا کرد، مانند: هم‌میهن، هم‌ارزی.
۸. ضمیرهای متصل چسبیده به کلمه‌ی پیش از خود نوشته می‌شوند، مانند: کتابم، نامت، کلامشان.

فصل ۳

کارهای پیشین

در فصل سوم پایان نامه، کارهای پیشین انجام شده روی مساله به تفصیل توضیح داده می شود. نمونه ای از فصل کارهای پیشین^۱ در زیر آمده است.

۱.۳ مساله های خوشه بندی

مساله ی خوشه بندی^۲ یکی از مهم ترین مسائل در زمینه ی داده کاوی به حساب می آید. در این مساله، هدف دسته بندی تعدادی شی به گونه ای است که اشیاء درون یک دسته (خوشه)، نسبت به یکدیگر در برابر دسته های دیگر شبیه تر باشند (معیارهای متفاوتی برای تشابه تعریف می گردد). این مساله در حوزه های مختلفی از علوم کامپیوتر از جمله داده کاوی، جست و جوی الگو^۳ پردازش تصویر^۴، بازیابی اطلاعات^۵ و رایانش زیستی^۶ مورد استفاده قرار می گیرد [۳]. تا کنون راه حل های زیادی برای این مساله ارایه شده است که از لحاظ معیار تشخیص خوشه ها و نحوه ی انتخاب یک خوشه، با یک دیگر تفاوت بسیاری دارند. به همین خاطر مساله ی خوشه بندی یک مساله ی بهینه سازی چندهدفه^۷ محسوب می شود.

همان طور که در مرجع [۴] ذکر شده است، خوشه در خوشه بندی تعریف واحدی ندارد و یکی از دلایل وجود الگوریتم های متفاوت، همین تفاوت تعریف ها از خوشه است. بنابراین با توجه به مدلی که برای خوشه ها ارایه می شود، الگوریتم متفاوتی نیز ارایه می گردد. در ادامه به بررسی تعدادی از معروف ترین مدل های مطرح می پردازیم.

- مدل های مرکزگرا: در این مدل ها، هر دسته با یک مرکز نشان داده می شود. از جمله معروف ترین روش های خوشه بندی بر اساس این مدل، خوشه بندی k -مرکز، خوشه بندی k -میانگین^۸ و خوشه بندی k -میان^۹ است.
- مدل های مبتنی بر توزیع نقاط: در این مدل ها، دسته ها با فرض پیروی از یک توزیع احتمالی مشخص می شوند. از جمله الگوریتم های معروف ارایه شده در این مدل، الگوریتم بیشینه سازی امید ریاضی^{۱۰} است.

^۱مطالب این فصل نمونه از پایان نامه ی آقای بهنام حاتمی گرفته شده است.

^۲Clustering

^۳Pattern Recognition

^۴Image Analysis

^۵Information Retrieval

^۶Bioinformatics

^۷Multi-Objective Optimization

^۸ k -Means

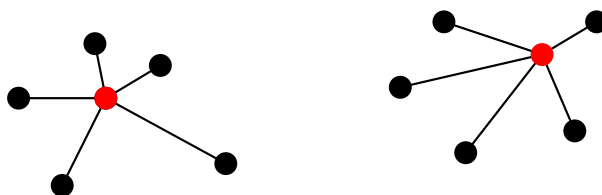
^۹ k -Median

^{۱۰}Expectation Maximization

- مدل‌های مبتنی بر تراکم نقاط: در این مدل‌ها، خوشه‌ها متناسب با ناحیه‌های متراکم نقاط در مجموعه داده مورد استفاده قرار می‌گیرد.
- مدل‌های مبتنی بر گراف: در این مدل‌ها، هر خوشه به مجموعه از راس‌ها گفته می‌شود که تمام راس‌های آن با یک‌دیگر همسایه باشند. از جمله الگوریتم‌های معروف این مدل، الگوریتم خوشه‌بندی HCS^۱ است. الگوریتم‌های ارایه شده تنها از نظر نوع مدل با یک‌دیگر متفاوت نیستند بلکه، می‌توان آن‌ها را از لحاظ نحوه‌ی تخصیص نقاط بین خوشه‌ها نیز تقسیم‌بندی کرد.
- تخصیص قطعی داده‌ها: در این نوع خوشه‌بندی هر داده دقیقاً به یک خوشه اختصاص داده می‌شود.
- تخصیص قطعی داده‌ها با داده‌ی پرت: در این نوع خوشه‌بندی ممکن است بعضی از داده‌ها به هیچ خوشه‌ای اختصاص نیابد، اما بقیه داده‌ها هر کدام دقیقاً به یک خوشه اختصاص می‌یابد.
- تخصیص قطعی داده: در این نوع خوشه‌بندی هر داده دقیقاً به یک خوشه اختصاص داده می‌شود.
- خوشه‌بندی هم‌پوشان: در این نوع خوشه‌بندی هر داده می‌تواند به چند خوشه اختصاص داده شود. در گونه‌ای از این مدل، می‌توان هر نقطه را با احتمالی به هر خوشه اختصاص می‌یابد. به این گونه از خوشه‌بندی، خوشه‌بندی نرم^۲ گفته می‌شود.
- خوشه‌بندی سلسه‌مراتبی: در این نوع خوشه‌ها، داده‌ها به گونه‌ای به خوشه‌ها تخصیص داده می‌شود که دو خوشه یا اشتراک ندارند یا یکی به طور کامل دیگری را می‌پوشاند. در واقع در بین خوشه‌ها، رابطه‌ی پدر فرزندی برقرار است. در بین دسته‌بندی‌های ذکر شده، تمرکز اصلی این پایان‌نامه بر روی مدل مرکزگرا و خوشه‌بندی قطعی با داده‌های پرت با مدل k -مرکز است. همان‌طور که ذکر شد علاوه بر مساله‌ی k -مرکز که به تفصیل مورد بررسی قرار می‌گیرد، k -میانه و k -میانگین از جمله معروف‌ترین خوشه‌بندی‌های مدل مرکزگرا هستند. در خوشه‌بندی k -میانه، هدف افزایش نقاط به k خوشه است به گونه‌ای که مجموع مربع فاصله‌ی هر نقطه از میانه‌ی نقاط آن خوشه، کمینه گردد. در خوشه‌بندی k -میانگین، هدف افزایش نقاط به k خوشه است به گونه‌ای که مجموع فاصله‌ی هر نقطه از میانگین نقاط داخل خوشه (یا مرکز آن خوشه) کمینه گردد.

۲.۳ خوشه‌بندی k -مرکز

یکی از رویکردهای شناخته‌شده برای مساله‌ی خوشه‌بندی، مساله‌ی k -مرکز است. در این مساله هدف، پیدا کردن k نقطه به عنوان مرکز دسته‌ها است به طوری که شعاع دسته‌ها تا حد ممکن کمینه شود. مثالی از مساله‌ی ۲-مرکز در شکل ۲.۳ نشان داده شده است. در این پژوهش، مساله‌ی k -مرکز با متریک‌های خاص و برای k های کوچک مورد بررسی قرار گرفته است و هر کدام از تعریف رسمی مساله‌ی k -مرکز در زیر آمده است.



شکل ۲.۳. نمونه‌ای از مساله‌ی ۲-مرکز

^۱Highly Connected Subgraphs

^۲Soft Clustering

مسئله ۱.۳ (k -مرکز). گراف کامل بدون جهت $G = (V, E)$ با تابع فاصله‌ی d ، که از نامساوی مثلثی پیروی می‌کند داده شده است. زیرمجموعه‌ی $S \subseteq V$ با اندازه‌ی k را به گونه‌ای انتخاب کنید که عبارت زیر را کمینه کند.

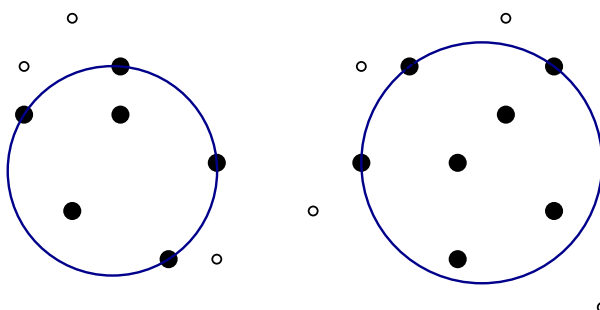
$$\max_{v \in V} \min_{s \in S} d(v, s) \quad (۱.۳)$$

گونه‌های مختلفی از مساله‌ی k -مرکز با محدودیت‌های متفاوت توسط پژوهشگران مورد مطالعه قرار گرفته است. از جمله‌ی این گونه‌ها، می‌توان به حالتی که در بین داده‌های ورودی، داده‌های پرت وجود دارد، اشاره کرد. در واقع در این مساله، قبل از خوشه‌بندی می‌توانیم تعدادی از نقاط ورودی را حذف نموده و سپس به خوشه‌بندی نقاط پردازیم. سختی این مساله از آنجاست که نه تنها باید مساله‌ی خوشه‌بندی را حل نمود، بلکه در ابتدا باید تصمیم گرفت که کدام یک از داده‌ها را به عنوان داده‌ی پرت در نظر گرفت که بهترین جواب در زمان خوشه‌بندی به دست آید. در واقع اگر تعداد نقاط پرتی که مجاز به حذف است، برابر صفر باشد، مساله به مساله‌ی k -مرکز تبدیل می‌شود. نمونه‌ای از مساله‌ی ۲-مرکز با ۷ داده‌ی پرت را در شکل ۴.۳ می‌توانید ببینید. تعریف دقیق‌ترین مساله در زیر آمده است.

مسئله ۲.۳ (k -مرکز با داده‌های پرت). یک گراف کامل بدون جهت $G = (V, E)$ با تابع فاصله‌ی d ، که از نامساوی مثلثی پیروی می‌کند داده شده است. زیرمجموعه‌ی $Z \subseteq V$ با اندازه‌ی z و مجموعه‌ی $S \subseteq V - Z$ با اندازه‌ی k را انتخاب کنید به طوری که عبارت زیر را کمینه کند.

$$\max_{v \in V-Z} \min_{s \in S} d(v, s) \quad (۲.۳)$$

گونه‌ی دیگری از مساله‌ی k -مرکز که در سال‌های اخیر مورد توجه قرار گرفته است، حالت جویبار داده‌ی آن است. در این گونه از مساله‌ی k -مرکز، در ابتدا تمام نقاط در دسترس نیستند، بلکه به مرور زمان نقاط در دسترس قرار می‌گیرند. محدودیت دومی که وجود دارد، محدودیت حافظه است، به طوری که نمی‌توان تمام نقاط را در حافظه نگه داشت و بعضاً حتی امکان نگه‌داری در حافظه‌ی جانبی نیز وجود ندارد و به‌طور معمول باید مرتبه‌ی حافظه‌ای کم‌تر از مرتبه حافظه‌ی خطی^۱ متناسب با تعداد نقاط استفاده نمود. از این به بعد به چنین مرتبه‌ای، مرتبه‌ی زیرخطی^۲ می‌گوییم. مدلی که ما در این پژوهش بر روی آن تمرکز داریم مدل جویبار داده تک‌گذره^۳ [۵] است. یعنی تنها یک بار می‌توان از ابتدا تا انتهای داده‌ها را بررسی کرد و پس از عبور از یک داده، اگر آن داده در حافظه ذخیره نشده باشد، دیگر به آن دسترسی وجود ندارد. علاوه بر این، در هر لحظه باید بتوان به پرسمان (برای تمام نقاطی از جویبار داده که تاکنون به آن دسترسی داشته‌ایم) پاسخ داد.



شکل ۴.۳. نمونه‌ای از مساله‌ی ۲-مرکز با داده‌های پرت

^۱Linear

^۲Sublinear

^۳Single pass

مساله ۳.۳ (k -مرکز در حالت جویبار داده). مجموعه‌ای از نقاط در فضای d -بعدی به مرور زمان داده می‌شود. در هر لحظه از زمان، به ازای مجموعه‌ی U از نقاطی که تا کنون وارد شده‌اند، زیرمجموعه‌ی $S \subseteq U$ با اندازه‌ی k را انتخاب کنید به طوری که عبارت زیر کمینه شود.

$$\max_{u \in U} \min_{s \in S} d(u, s) \quad (۳.۳)$$

از آنجایی که گونه‌ی جویبار داده و داده پرت مساله‌ی k -مرکز به علت به‌روز بودن مبحث داده‌های حجیم^۱، به تازگی مورد توجه قرار گرفته است. در این تحقیق سعی شده است که تمرکز بر روی این گونه‌ی خاص از مساله باشد. همچنین در این پژوهش سعی می‌شود گونه‌های مساله را برای انواع متریک‌ها و برای k های کوچک نیز مورد بررسی قرار داد.

۳.۳ مدل جویبار داده

همان‌طور که ذکر شد مساله‌ی k -مرکز در حالت داده‌های پرت و جویبار داده، گونه‌های تعمیم‌یافته از مساله‌ی k -مرکز هستند و در حالت‌های خاص به مساله‌ی k -مرکز کاهش پیدا می‌کنند. مساله‌ی k -مرکز در حوزه‌ی مسائل ان‌پی-سخت^۲ قرار می‌گیرد و با فرض $P \neq NP$ الگوریتم دقیق با زمان چندجمله‌ای برای آن وجود ندارد [۶]. بنابراین برای حل کارای^۳ این مسائل از الگوریتم‌های تقریبی^۴ استفاده می‌شود.

برای مساله‌ی k -مرکز، دو الگوریتم تقریبی معروف وجود دارد. در الگوریتم اول، که به روش حریصانه^۵ عمل می‌کند، در هر مرحله بهترین مرکز ممکن را انتخاب می‌کند به طوری تا حد ممکن از مراکز قبلی دور باشد [۷]. این الگوریتم، الگوریتم تقریبی با ضریب تقریب ۲ ارایه می‌دهد. در الگوریتم دوم، با استفاده از مساله‌ی مجموعه‌ی غالب کمینه^۶، الگوریتمی با ضریب تقریب ۲ ارایه می‌گردد [۸]. همچنین ثابت شده است، که بهتر از این ضریب تقریب، الگوریتمی نمی‌توان ارایه داد مگر آن که $P = NP$ باشد.

برای مساله‌ی k -مرکز در حالت جویبار داده برای ابعاد بالا، بهترین الگوریتم موجود ضریب تقریب $2 + \varepsilon$ دارد [۹، ۱۰، ۱۱] و ثابت می‌شود الگوریتمی با ضریب تقریب بهتر از ۲ نمی‌توان ارایه داد. برای مساله‌ی k -مرکز با داده‌ی پرت در حالت جویبار داده نیز، بهترین الگوریتم ارایه شده، الگوریتمی با ضریب تقریب $4 + \varepsilon$ است که با کران پایین ۳ هنوز اختلاف قابل توجهی دارد [۱۲].

برای k های کوچ به خصوص $k = 1, 2$ ، الگوریتم‌های بهتری ارایه شده است. بهترین الگوریتم ارایه شده برای مساله‌ی ۱-مرکز در حالت جویبار داده برای ابعاد بالا، دارای ضریب تقریب $1/22$ است و کران پایین $\frac{1+\sqrt{2}}{4}$ نیز برای این مساله اثبات شده است [۱۳، ۱۴]. برای مساله ۲-مرکز در حالت جویبار داده برای ابعاد بالا، اخیراً راه‌حلی با ضریب تقریب $1/8 + \varepsilon$ ارایه شده است [۱۵]. برای مساله‌ی ۱-مرکز با داده‌ی پرت، تنها الگوریتم موجود، الگوریتمی با ضریب تقریب $1/73$ است [۱۶].

^۱Big Data

^۲NP-Hard

^۳Efficient

^۴Approximation Algorithm

^۵Greedy

^۶Dominating Set

۴.۳ تقریب‌پذیری

یکی از راه‌کارهایی که برای کارآمد کردن راه‌حل ارایه شده برای یک مساله وجود دارد، استفاده از الگوریتم‌های تقریبی برای حل آن مساله است. یکی از عمده‌ترین دغدغه‌های مطرح در الگوریتم‌های تقریبی کاهش ضریب تقریب است. در بعضی از موارد حتی امکان ارایه‌ی الگوریتم تقریبی با ضریبی ثابت نیز وجود ندارد. به طور مثال، الگوریتم تقریبی با ضریب تقریب کم‌تر از ۲، برای مساله‌ی k -مرکز وجود ندارد مگر این‌که $P = NP$ باشد. برای مسائل مختلف، معمولاً می‌توان کران پایینی برای میزان تقریب‌پذیری آن‌ها ارایه داد. در واقع برای برخی مسائل ان‌پی-سخت، علاوه بر این که الگوریتم کارآمدی وجود ندارد، گاهی الگوریتم تقریبی با ضریبی تقریب کم و نزدیک به یک نیز وجود ندارد. در جدول ۲.۳ میزان تقریب‌پذیری مسائل مختلفی که در این پایان‌نامه مورد استفاده قرار می‌گیرد را می‌بینید.

جدول ۲.۳. نمونه‌هایی از کران پایین تقریب‌پذیری مسائل خوشه‌بندی

مساله	کران پایین تقریب‌پذیری	مرجع
k -مرکز	۲	[۸]
k -مرکز در فضای اقلیدسی	$1/822$	[۱۷]
۱-مرکز در حالت جویبار داده	$\frac{1+\sqrt{2}}{4}$	[۱۳]
k -مرکز با نقاط پرت و نقاط اجباری	۳	[۱۲]

فصل ۴

نتایج جدید

در این فصل نتایج جدید به دست آمده در پایان نامه توضیح داده می شود. در صورت نیاز می توان نتایج جدید را در قالب چند فصل ارایه نمود. همچنین در صورت وجود پیاده سازی، بهتر است نتایج پیاده سازی را در فصل مستقلی پس از این فصل قرار داد.

فصل ۵

نتیجه‌گیری

در این فصل، ضمن جمع‌بندی نتایج جدید ارائه‌شده در پایان‌نامه یا رساله، مساله‌های باز باقی‌مانده و همچنین پیشنهادهایی برای ادامه‌ی کار ارائه می‌شوند.

پیوست آ

مطالب تکمیلی

پیوست‌های خود را در صورت وجود می‌توانید در این قسمت قرار دهید.

- [1] D. E. Knuth. *The T_EXbook*. Addison-Wesley, 1984.
- [2] L. Lamport. *L_AT_EX—A Document Preparation System*. Addison-Wesley, 1985.
- [3] J. Han and M. Kamber. *Data Mining, Southeast Asia Edition: Concepts and Techniques*. Morgan kaufmann, 2006.
- [4] V. Estivill-Castro, “Why so many clustering algorithms: a position paper,” *ACM SIGKDD explorations newsletter*, vol.4, no.1, pp.65–75, 2002.
- [5] C. C. Aggarwal. *Data streams: models and algorithms*. Springer Science & Business Media, 2007.
- [6] M. R. Garey and D. S. Johnson, “Computers and intractability: a guide to the theory of NP-completeness,” *Freeman & Co.*, 1979.
- [7] N. Megiddo and K. J. Supowit, “On the complexity of some common geometric location problems,” *SIAM J. Computing*, vol.13, no.1, pp.182–196, 1984.
- [8] V. V. Vazirani. *Approximation Algorithms*. Springer-Verlag New York, Inc., 2001.
- [9] R. M. McCutchen and S. Khuller, “Streaming algorithms for k-center clustering with outliers and with anonymity,” pp.165–178, 2008.
- [10] S. Guha, “Tight results for clustering and summarizing data streams,” pp.268–275, 2009.
- [11] H.-K. Ahn, H.-S. Kim, S.-S. Kim, and W. Son, “Computing k centers over streaming data for small k,” vol.24, no.02, pp.107–123, 2014.
- [12] M. Charikar, S. Khuller, D. M. Mount, and G. Narasimhan, “Algorithms for facility location problems with outliers,” pp.642–651, 2001.
- [13] P. K. Agarwal and R. Sharathkumar, “Streaming algorithms for extent problems in high dimensions,” pp.1481–1489, 2010.
- [14] T. M. Chan and V. Pathak, “Streaming and dynamic algorithms for minimum enclosing balls in high dimensions,” vol.47, no.2, pp.240–247, 2014.
- [15] S.-S. Kim and H.-K. Ahn, “An improved data stream algorithm for clustering,” pp.273–284, 2014.
- [16] H. Zarrabi-Zadeh and A. Mukhopadhyay, “Streaming 1-center with outliers in high dimensions,” pp.83–86, 2009.
- [17] M. Bern and D. Eppstein, “Approximation algorithms for geometric problems,” in *Approximation Algorithms for NP-hard Problems* (D. S. Hochbaum, ed.), pp.296–345, PWS Publishing Co., 1997.

واژه‌نامه

extreme..... حدی	greedy حرصانه	heuristic ابتکاری	high dimensions ابعاد بالا
cluster خوشه	linear خطی	bias اربب	threshold آستانه
		pigeonhole principle اصل لانه‌ی کبوتری	NP-Hard ان‌پی-سخت
		transition انتقال	
data داده	data mining داده‌کاوی	online برخط	linear programming برنامه‌ریزی خطی
outlier data داده‌ی پرت	doubling دوبرابر سازی	optimum بهینه	maximum بیشینه
binary دودویی			
vertex راس	formal رسمی	outlier پرت	query پرسمان
		cover پوشش	complexity پیچیدگی
sublinear زیرخطی			
amortized سرشکن	hierarchichal سلسه‌مراتبی	experimental تجربی	density تراکم
pseudocode شبه کد	object شی	approximation تقریب	partition تقسیم‌بندی
		mesh توری	distributed توزیع شده
satisfiability صدق‌پذیری		separable جداپذیر	black box جعبه سیاه
dominate غلبه		data stream جویبار داده	

parallelization موازی سازی
buffer میان گیر

ن

inversion نابه جایی
invariant ناورد
center point نقطه ی مرکزی
half space نیم فضا

ه

price of anarchy (POA) هزینه ی آشوب

ی

edge یال

ف

distance فاصله
space فضا

ق

deterministic قطعی

ک

efficient کارا
candidate کاندیدا
minimum کمینه

م

set مجموعه
coreset مجموعه هسته
planar مسطح

Abstract

We present a standard template for typesetting theses in Persian. The template is based on the X_{P} Persian package for the \LaTeX typesetting system. This write-up shows a sample usage of this template.

Keywords. Thesis, Typesetting, Template, X_{P} Persian.



Sharif University of Technology
Department of Mathematical Sciences

Master's Thesis
Computer Science

A Standard Template for Typesetting Theses in Persian

Written by
The Author

Supervisor
Your Supervisor

September 2025