



دانشگاه تبریز

دانشگاه تبریز

دانشکده ریاضی

گروه علوم کامپیوتر

پایان نامه

برای دریافت درجه کارشناسی ارشد در رشته علوم کامپیوتر

عنوان

دسته بندی صفحات وب برای پالایش

استاد راهنما:

دکتر محمد رضا فیضی درخشی

استاد مشاور:

محمد علی بالافر

پژوهشگر:

شیمای رشیدی

شهریور ۱۳۹۲

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

ماحصل آموخته‌هایم را تقدیم می‌کنم به آنان که مهر آسمانی‌شان آرام‌بخش آلام زمینی‌ام است

به استوارترین تکیه‌گاه، دستان پر مهر پدرم

و به دلسوزترین نگاه، چشمان شورانگیز مادرم

که هرچه آموختم در مکتب عشق شماست و هرچه بلو شدم قطره‌ای از دریای بی‌کران عشق‌تان را پاس‌گزار نیستم

باشد که حاصل تلاشم نسیم کوزه‌غبار حسگیتان را بروداید.

سپاس‌گزاری...

سپاس بی‌کران پروردگاریکتار که هستی‌مان بخشید و به طریق علم و دانش را بنمونمان شد.

بر خود لازم می‌دانم از کلیه کسانی که بنده را در تدوین و نگارش این پایان‌نامه یاری نمودند صمیمانه تشکر و قدردانی نمایم؛ به خصوص از استاد

راهبهای فرزانه خود جناب آقای دکتر محمد رضا فیضی در خشی که در کلیه مراحل انجام این پژوهش با خوشرویی، یاری و راهبهای ام نمودند. همچنین

از استاد فرهیخته جناب آقای دکتر محمد علی بالافکر که وقت خود را بی‌شائبه در اختیار اینجانب گذاشته و با دقت نظر خاصی مشاوره لازم را در این

خصوص ارائه نمودند کمال امتنان را دارم. از استاد صبور جناب آقای دکتر شهریار لطفی، مدیرت محترم گروه که بدون مساعدت ایشان،

این پروژه به نتیجه مطلوب نمی‌رسید و همچنین زحمت داورسی این پایان‌نامه را متقبل فرمودند کمال تشکر را دارم.

همچنین از جناب آقای دکتر نظری استاد که اقدر گروه روانشناسی و آقای دکتر جهان که با کمک های ایشان خود مرا در انجام این پروژه

یاری نمودند و دیگر اساتید ارجمند که در محضرشان کسب علم نمودم تقدیر و قدردانی می‌نمایم.

از تمامی دوستان و عزیزانی که در طول انجام این تحقیق از یاری‌شان برخوردار بودم تشکر نموده و برای تمامی این عزیزان خواستار

سلامت و موفقیت از درگاه ایزدمنان، هستم.

شماره‌ی

شهریور ۹۲

نام خانوادگی دانشجو: رشیدی	نام: شیما
عنوان پایان نامه: دسته‌بندی صفحات وب برای پالایش	
استاد راهنما: دکتر محمد رضا فیضی درخشی	
استاد مشاور: دکتر محمد علی بالافر	
مقطع تحصیلی: کارشناسی ارشد	رشته: علوم کامپیوتر
دانشگاه: دانشگاه تبریز	دانشکده: علوم ریاضی
تاریخ فارغ التحصیلی: شهریور ۱۳۹۲	تعداد صفحه: ۷۰
کلید واژه‌ها: طبقه‌بندی، پالایش صفحات وب، فراوانی کلمات و همنشینی کلمات	
چکیده:	
<p>مسئله حفاظت از اқشار آسیب‌پذیر جامعه در مقابل موارد ناهنجار در صفحات وب، موجب احساس نیاز به پالایش صفحات وب شده است. یکی از جنبه‌های پالایش صفحات وب که امروزه بیش از پیش مورد توجه قرار گرفته است، طبقه‌بندی صفحات بر اساس میزان خشونت‌آمیز بودن آن‌ها است. طبقه‌بندی در واقع نسبت دادن یک صفحه به یک یا چند موضوع از پیش تعیین شده با استفاده از الگوهای آموزش داده شده می‌باشد. سیستم با استفاده از یک سری صفحات آموزشی که موضوع آن‌ها از قبل تعیین شده است، آموزش داده می‌شود تا بتواند اسناد جدید را به صورت صحیح طبقه‌بندی نماید. روش‌های مختلفی برای این کار وجود دارد که اغلب روش‌های آماری یا هوش مصنوعی می‌باشند.</p> <p>در این پایان‌نامه رویکرد همنشینی کلمات برای طبقه‌بندی صفحات وب فارسی پیشنهاد شده است. در واقع با استخراج میزان همنشینی کلمات خشونت‌آمیز با هم و با کلمات تعدیل‌کننده خشونت، متون فارسی به دو دسته خشونت‌آمیز و غیرخشونت‌آمیز طبقه‌بندی می‌شوند. به دلیل عدم وجود مجموعه داده استاندارد، صفحات وب به صورت دستی جمع‌آوری شده است. در روش ارائه شده، در فاز آموزش از ۷۰ داده آموزشی برچسب‌دار استفاده شده است. در این فاز با استفاده از داده‌ها، میزان همنشینی کلمات خشونت‌آمیز با هم و با کلمات تعدیل‌کننده خشونت محاسبه شده و ماشین آمادۀ طبقه‌بندی صفحات جدید می‌شود. برای سنجش کیفیت کار نیز از ۱۲۰ صفحه وب استفاده شده است و در نهایت الگوریتم طبقه‌بندی‌کننده توانسته است به دقت ۹۷/۱۷۵٪ دست پیدا کند.</p>	

فهرست

شماره صفحه	عنوان
۱	فصل اول: مقدمه
۲	۱-۱ مقدمه
۲	۲-۱ شرح مسئله
۳	۳-۱ ساختار پایان نامه
۵	فصل دوم: مفاهیم پایه
۶	۱-۲ مقدمه
۶	۲-۲ پالایش در اینترنت
۷	۱-۲-۲ مکان استقرار پالایش
۹	۲-۲-۲ لایه‌های عملیاتی پالایش
۱۲	۳-۲-۲ روش‌های پالایش
۱۵	۴-۲-۲ پالایش در کشورهای مختلف
۱۸	۳-۲ طبقه‌بندی و روش‌های آن
۱۹	۱-۲-۲ درخت‌های تصمیم
۲۲	۲-۲-۲ Naive Bayes
۲۳	۳-۲-۲ k همسایه نزدیک
۲۴	۴-۲-۲ شبکه‌های عصبی
۲۵	۵-۲-۲ ماشین بردار پشتیبان
۲۶	۴-۲ نمایش سند و ارزیابی طبقه‌بندی کننده
۲۶	۱-۴-۲ نمایش سند
۲۸	۲-۴-۲ پیش پردازش متن
۳۱	۳-۴-۲ معیارهای ارزیابی
۳۳	۵-۲ جمع‌بندی فصل
۳۵	فصل سوم: بررسی روش‌های موجود
۳۶	۱-۳ مقدمه
۳۶	۲-۳ انطباق کلید واژه
۳۷	۳-۳ تشابه اسناد

۴۱	۴-۳ فراوانی کلمات
۴۵	۵-۳ جمع‌بندی فصل
۴۶	فصل چهارم: طبقه‌بندی صفحات وب فارسی با رویکرد همنشینی کلمات
۴۷	۱-۴ مقدمه
۴۷	۲-۴ انتخاب ابزار
۴۷	۳-۴ مجموعه داده
۴۸	۴-۴ پیش پردازش متن
۴۹	۱-۴-۴ نمایش سند
۵۰	۵-۴ طبقه‌بندی صفحات وب با رویکرد همنشینی کلمات
۵۳	۱-۵-۴ نتایج آزمایش
۵۶	۶-۴ جمع‌بندی فصل
۵۷	فصل پنجم: نتیجه‌گیری و پیشنهادهای آینده
۵۸	۱-۵ نتیجه‌گیری
۶۰	۲-۵ پیشنهادهای آینده
۶۲	مراجع
۶۰	پیوست

فهرست شکل‌ها

عنوان	شماره صفحه
شکل ۱-۲: محل قرارگیری فیلتر [۳]	۹
شکل ۲-۲: روش‌های فیلترینگ	۱۲
شکل ۳-۲: مدل ریاضی ساده شده عصب واقعی [۱۲]	۲۴
شکل ۱-۴: نحوه نمایش یک صفحه با ماتریس‌های همنشینی کلمات	۵۰

فهرست جداول

شماره صفحه	عنوان
۱۱	جدول ۱-۲: مدل OSI [۳]
۳۲	جدول ۲-۲: جدول حالات ممکن برای نتیجه یک طبقه‌بندی کننده
۳۲	جدول ۳-۲: فرمول‌های سنجش کارایی طبقه‌بندی کننده‌های باینری
۴۰	جدول ۲-۳: نتایج آزمایش برای ۱۰۰ بار اجرا [۱۷]
۴۳	جدول ۳-۳: ویژگی‌های مورد استفاده برای طبقه‌بندی صفحات وب [۲۰]
۵۳	جدول ۱-۴: تعداد صفحات وب مجموعه آموزشی و آزمایشی
۵۴	جدول ۲-۴: نتایج حاصل از رویکرد ارائه شده
۵۴	جدول ۳-۴: مقادیر FP و TP برای آزمایش‌های انجام شده
۵۵	جدول ۴-۴: مقایسه کارایی روش‌های طبقه‌بندی صفحات وب جهت فیلترینگ
۵۶	جدول ۵-۴: زمان مورد نیاز برای طبقه‌بندی

فصل اول:

مقدمه

۱-۱ مقدمه

پالایش^۱ صفحات وب یکی از روش‌های کنترل محتوای اینترنت است که امروزه بسیار مورد توجه می‌باشد. از آنجایی که با توسعه سریع شبکه جهانی وب، مردم روز به روز بیشتر از اشتراک‌گذاری اطلاعات بهره می‌گیرند، تنوع دیدگاه‌ها و توانایی برای ارسال هر نوع اطلاعاتی روی اینترنت، موجب شده است تا دسترسی به محتوای موهن، مضر و غیرقانونی بیش از پیش آسان گردد. از این رو پالایش صفحات وب حائز اهمیت می‌باشد که از جنبه‌های مورد توجه در این زمینه، شناسایی و طبقه‌بندی^۲ صفحات بر اساس میزان خشونت‌آمیز بودن آن‌هاست. با طبقه‌بندی صفحات وب می‌توان چنین صفحاتی را از صفحات سالم تشخیص داده و از دسترسی آسان به این‌گونه مطالب جلوگیری کرد. طبقه‌بندی صفحه وب، در واقع اختصاص یک صفحه وب به یک یا چند برچسب کلاس از پیش تعریف شده است. طبقه‌بندی اغلب به عنوان یک مسئله یادگیری نظارت شده مطرح می‌شود که در آن مجموعه‌ای از داده‌های برچسب‌دار برای آموزش طبقه‌بندی‌کننده استفاده می‌شوند. سپس طبقه‌بندی‌کننده می‌تواند به برچسب نمونه‌های آینده اعمال شود. در این فصل به شرح مسئله و ساختار پایان‌نامه پرداخته می‌شود.

۱-۲ شرح مسئله

در بیش از یک دهه گذشته وب رسانه‌ای بسیار پرکاربرد برای انتشار اطلاعات بوده است. رشد و گسترش شبکه جهانی وب، دسترس‌پذیری به همه اشکال مختلف اطلاعاتی اعم از متن و فایل‌های صوتی، تصویری و گرافیکی متنوع را برای کاربران به راحتی میسر کرده است. در این میان انتشار برخی اطلاعات با محتوای غیراخلاقی، ترویج خشونت، تجارت مواد مخدر و موارد مشابه که دسترسی به آن‌ها در شبکه اینترنت به آسانی صورت می‌گیرد [۱]، آثار مخرب بسیاری بر افراد به خصوص کودکان می‌گذارد. این مسئله دولت‌ها و مقامات مسئول را بر آن داشته است تا کنترل و نظارت بیشتری تا حد مسدود کردن دسترسی به پایگاه‌های مخرب و برخورد قضایی و قانونی با گردانندگان آن‌ها اعمال نمایند که در اصطلاح به آن «فیلترینگ» گفته می‌شود. این موضوع با توجه به نگرش‌های خاص فرهنگی و اجتماعی حاکم بر فرهنگ جوامع مختلف متفاوت است [۱]. از آنجایی که طبقه‌بندی نقشی حیاتی در مدیریت و بازیابی

^۱ filtering

^۲ classification

اطلاعات دارد و به بهبود کیفیت جستجو کمک می‌کند [۲]، می‌توان از روش‌های طبقه‌بندی صفحات وب و ویژگی‌های آن‌ها برای پالایش استفاده کرد.

طبقه‌بندی صفحات وب که به عنوان دسته‌بندی^۱ صفحات وب نیز شناخته می‌شود، فرآیند انتساب یک صفحه وب به یک یا چند دسته برچسب از پیش تعریف شده است. طبقه‌بندی صفحات وب شامل طبقه‌بندی صفحات بر اساس پارامترهای مختلف مانند متن، تصویر، ساختار سند و غیره است. طبقه‌بندی بر اساس تعداد کلاس‌ها می‌تواند به طبقه‌بندی باینری یا طبقه‌بندی چند کلاسه تقسیم شود.

هدف این پایان‌نامه شناسایی و طبقه‌بندی صفحات وب فارسی از لحاظ خشونت، با دقت مناسب می‌باشد. در واقع می‌خواهیم درصد پالایش صفحات وب خشونت‌آمیز را بالا برده و درصد پالایش صفحات وب غیرخشونت‌آمیز را که به اشتباه مضر شناسایی شده را کاهش دهیم. جهت رسیدن به این هدف ابتدا مفاهیم پایه‌ای در این زمینه بررسی خواهد شد. سپس روش‌های پالایش موجود، مورد ارزیابی قرار خواهد گرفت. در نهایت نیز روشی با دقت بالاتر برای طبقه‌بندی صفحات وب از لحاظ خشونت در زبان فارسی ارائه می‌شود.

۱-۳ ساختار پایان‌نامه

در ادامه بحث در فصل ۲، مسائل مربوط به پالایش، مفاهیم پایه‌ای که برای طبقه‌بندی صفحات وب با آن‌ها سروکار داریم، و همچنین نمایش سند و ارزیابی عملکرد طبقه‌بندی‌کننده توضیح داده شده است. ابتدا اشاره‌ای به مبحث پالایش شده است. سپس لایه‌های عملیاتی، محل استقرار پالایش و روش‌های پالایش بررسی و نحوه پالایش صفحات وب در کشورهای مختلف تشریح شده است. در ادامه فصل ۲ توضیحاتی در مورد طبقه‌بندی داده شده است و برخی از روش‌های طبقه‌بندی، از جمله درخت‌های تصمیم، Naive Bayes، k همسایه نزدیک، شبکه‌های عصبی، ماشین بردار پشتیبان به اختصار بیان و ویژگی‌های آن‌ها بررسی شده است. در انتها نیز به نمایش سند، پیش پردازش متن و معیارهای کارایی طبقه‌بندی‌کننده پرداخته شده است. در فصل ۳ تعدادی از راه‌کارهایی که در دسته راه‌کارهای مبتنی بر

^۱ categorization

متن می‌باشند، و قبلاً در زمینه پالایش صفحات وب ارائه شده‌اند بررسی و ارزیابی آن‌ها بیان شده است. در فصل ۴ روش ارائه شده در این پایان‌نامه برای طبقه‌بندی صفحات وب فارسی جهت پالایش از نظر خشونت، شرح داده شده است. در ادامه فصل ۴ به آزمایش‌های انجام شده جهت بررسی کارایی این روش و نتایج حاصل، پرداخته شده است. در نهایت در فصل ۵ نتیجه‌گیری و کارهای آینده ذکر شده است.

فصل دوم:

مغایم نام
" " " " " "

۲-۱ مقدمه

در این فصل ابتدا مفاهیم اولیه در زمینه فیلترینگ در اینترنت و روش‌های آن ارائه می‌شود. سپس تعریف طبقه‌بندی و روش‌های طبقه‌بندی مورد استفاده در فیلترینگ بیان می‌شود. در آخر نیز به نحوه نمایش سند و ارزیابی طبقه‌بندی‌کننده پرداخته می‌شود. هدف در این فصل درک کلی مسئله، بررسی انواع راه‌حل‌های ارائه شده برای مسئله فیلترینگ، آشنایی با روش‌های مختلف طبقه‌بندی، چگونگی نمایش سند و همچنین نحوه ارزیابی یک طبقه‌بندی‌کننده می‌باشد.

۲-۲ پالایش در اینترنت

گسترش اینترنت زمینه‌ای را به وجود آورده است، که طیف وسیعی از اطلاعات بدون هیچ‌گونه محدودیتی و فراتر از مرزهای جغرافیایی در سراسر جهان منتشر شود و به نحو چشمگیری به یک رسانه ارتباطی و اطلاعاتی تبدیل گردد. ضمن اینکه مخاطبان و کاربران اینترنت هم به همان اندازه افزایش مداومی داشته‌اند. خصوصیت و ویژگی اینترنت و سهولت انتشار مطالب و اطلاعات و علاوه بر آن راحتی دستیابی به انواع مختلف اطلاعات در وب اعم از متن، صوت و تصویر و داده‌های گرافیکی متنوع، به اندازه‌ای شدت یافته است که در برخی موارد اطلاعات حاوی مطالب مخرب و زیانباری نیز در وب منتشر می‌شود که زمینه سوء استفاده‌های مختلف را فراهم آورده است. تصاویر غیراخلاقی به خصوص در زمینه سوء استفاده از کودکان، ترویج خشونت و فساد، خرید و فروش مواد مخدر و زیانبار، اشاعه اطلاعات خصوصی افراد و سازمان‌ها، استفاده از اطلاعات وب در جهت اهداف تروریستی و موارد مشابه، ضرورت کنترل محتوای وب را ناگزیر ساخته است. این مسئله دولت‌ها و مقامات مسئول را بر آن داشته است تا کنترل و نظارت بیشتری تا حد مسدود کردن دسترسی به پایگاه‌های مخرب و برخورد قضایی و قانونی با گردانندگان آن‌ها اعمال نمایند که در اصطلاح به آن «فیلترینگ»^۱ گفته می‌شود.

به تعبیری می‌توان گفت فیلترینگ اینترنتی یا فیلترینگ، عبارت است از محدود کردن دسترسی کاربران اینترنت به وبگاه‌ها و خدمات اینترنتی که از دیدگاه متولیان فرهنگی و سیاسی هر کشور برای

^۱ پالایش

مصرف عموم مناسب نیست، اعمال فیلتر به وسیله ارائه‌دهندگان خدمات اینترنتی انجام می‌شود ولی تعیین سطح، مصادیق و سیاست‌های فیلترینگ با حکومت‌هاست.

این کنترل و نظارت بر انتشار اطلاعات در جوامع مختلف و بر حسب خط‌مشی‌های سیاسی، نگرش‌های خاص فرهنگی و اجتماعی حاکم بر فرهنگ جوامع مختلف نموده‌های متفاوتی داشته است. در برخی جوامع غربی تأکید بیشتر بر عدم سوء استفاده از اطلاعات خصوصی افراد، جلوگیری از انتشار تصاویر جنسی کودکان، جلوگیری از ترویج خشونت و ممانعت از دسترسی سازمان‌های تروریستی به اطلاعات خاص و مهم است اما کنترل محتوای اینترنت در جوامع شرقی و مذهبی فراتر از موارد فوق است و نظارت شدید بر اشاعه افکار و عقاید سیاسی در شبکه جهانی را نیز شامل می‌شود.

دلایل عمده فیلتر کردن اینترنت در کشورهای مختلف را می‌توان در چهار دسته کلی گنجانند:

(۱) مسائل سیاسی

(۲) مسائل اجتماعی

(۳) مسائل امنیتی

(۴) مسائل اخلاقی

بحث نظارت و کنترل بر محتوای وب، چالش‌های سیاسی و فرهنگی را پدید آورده است که از جمله آن‌ها می‌توان به مخالفت‌های گسترده‌ای که با سانسور افکار و عقاید صورت می‌گیرد و نیز تأکید بر حق دسترسی کاربران به اطلاعات سالم و ارزشمند اشاره کرد. همان‌گونه که ذکر شد، این موضوع در هر کشوری بر حسب سیاست‌ها و ارزش‌های فرهنگی حاکم بر جامعه آن متفاوت است و ابزارهایی نیز که در زمینه کنترل و نظارت بر وب به کار گرفته می‌شوند، بسته به نوع مخاطبان و کاربران هر بخشی از جامعه، متفاوت و متنوع هستند.

۲-۲-۱ مکان استقرار پالایش

یک فیلتر وب ممکن است در سطوح مختلف از مدل OSI به کار گرفته شود و همانطور که در شکل ۲-۱ نشان داده شده است، می‌تواند در مکان‌های مختلفی در شبکه قرار گیرد. موارد یک تا پنج محل فیلتر را

مشخص می‌کنند. گزینه‌های سفارشی‌سازی، کارآیی فیلتر و امنیت فراهم شده تا حد زیادی به محل استقرار^۱ بستگی دارد. [۳]

۱. در سطح ملی/کشوری: فیلتر بین پیکره^۲ اینترنت و شبکه کشور مستقر شده است. کاربران هیچ کنترلی روی سفارشی‌سازی فیلتر ندارند، کارآیی آن و سپس امنیت فراهم شده با فیلتر توسط سیاست دولت تعیین شده است.

۲. در سطح سازمانی: فیلتر بین شبکه سازمان و گذرگاه اینترنت مستقر شده است. برای تمام کاربران این گذرگاه، مانند تمام کارکنان سازمان، فیلتر محتوای اینترنت فراهم شده است. فیلتر توسط مدیران وب سایت سازمان‌ها قابل تنظیم است که سیاست سازمان را با توجه به اینکه چه چیزی مناسب است و چه چیزی برای سازمان مناسب نیست، در نظر می‌گیرند. سازمان علاوه بر ارائه محتوای فیلتر شده نیز می‌تواند مدت زمان استفاده از اینترنت را محدود کند.

۳. در ارائه دهنده سرور اینترنت: فیلتر در گذرگاه ISP نصب شده و فیلتر محتوا را برای تمام کاربران خود فراهم می‌کند. فشارهای غیررسمی دولت در کانادا و انگلستان ISP‌های بزرگ را بر آن داشتند تا بطور داوطلبانه فیلترینگ مؤسسه‌ای برای مسدود کردن دسترسی به تصاویر غیراخلاقی کودکان و سوء استفاده‌های جنسی از کودکان انجام دهند. دادگاه‌های فرانسه، بلژیک و آلمان به ISP‌ها دستور داده‌اند تا سخنانی‌های منفور و اشتراک‌گذاری‌های نقطه به نقطه غیرقانونی فایل را مسدود کنند.

۴. در سطح فردی: فیلترینگ در کامپیوترهای محلی یا ایستگاه‌های کاری نصب شده است. فیلتر ممکن است قسمتی از فایروال، بسته آنتی ویروس باشد، یا از طریق برخی سیستم‌های مشابه دیگر مانند راهنما محتوا^۳، کنترل والدین^۴ و غیره صورت گیرد.

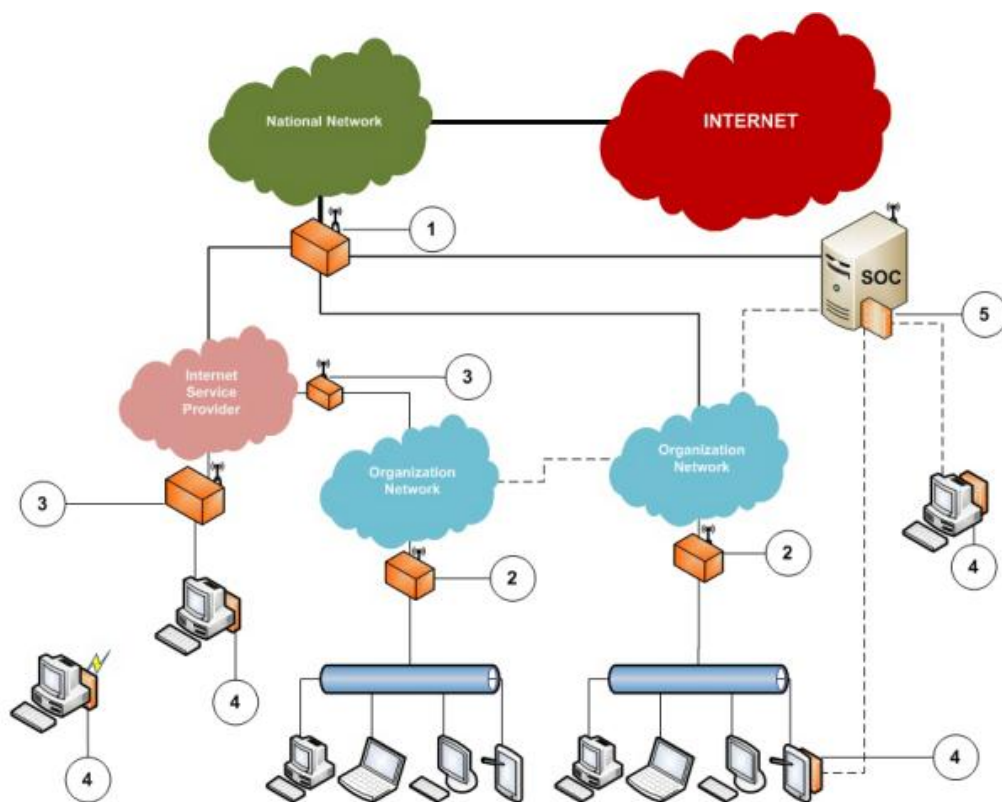
¹ deployment

² backbone

³ content advisor

⁴ parental control

۵. در سطح شخص ثالث^۱: خدمات فیلترینگ توسط یک فروشنده شخص ثالث مورد اعتماد از طریق مراکز عملیاتی امنیت (SOCs) خود آن فراهم می‌شود. مشتریان ترافیک وب خود را از طریق این SOCها با پروکسی کردن ارسال می‌کنند. ScanSafe و WebSense مثال‌هایی از فروشنده‌های شخص ثالث هستند. اگرچه برای همه نوع از سازمان‌ها و کاربران مناسب است، این سرویس به سازمان‌های کوچک و متوسط محدود شده است اما فیلترینگ را در هر سطح از یک سازمان ارائه می‌کند.



شکل ۲-۱: محل قرارگیری فیلتر [۳]

۲-۲-۲ لایه‌های عملیاتی پالایش

فیلتر نصب شده ممکن است در مکان‌های مختلف از مدل شبکه OSI، همانطور که در شکل ۲-۴ می‌بینید به کار گرفته شود: در لایه ۳ که شبکه نام دارد، در لایه ۴ که لایه انتقال نام دارد یا در لایه ۷ که لایه برنامه‌های کاربردی است. فیلتر نصب شده در لایه ۳ و ۴ به فیلترهای لایه شبکه اشاره دارند و آن‌هایی که در لایه ۷ نصب شده‌اند فیلترهای لایه کاربرد نامیده می‌شوند. [۳]

۱. لایه ۳ از مدل OSI مسئول آدرس‌دهی منطقی و مسیریابی داده‌ها با استفاده از پروتکل‌هایی مانند IP می‌باشد. بسته شامل آدرس‌های منبع و مقصد است که می‌تواند برای مسدود کردن انتقال بسته مبتنی بر برخی قوانین تعریف شده از فیلتر، استفاده شود.

۲. لایه ۴ از مدل OSI مسئول قالب‌بندی و حمل داده با استفاده از پروتکل‌هایی مانند TCP و UDP است. بسته‌ها در این لایه شامل آدرس‌های منبع و مقصد هستند. علاوه بر این حاوی اطلاعاتی در مورد نوع ترافیک شبکه می‌باشند در نتیجه انسداد ترافیک از آدرس خاص برای یک کاربرد خاص را مقدور می‌سازند.

۳. لایه ۷ از مدل OSI، مسئول تجزیه و تحلیل داده قبل از ارسال آن به برنامه کاربردی خاصی است. در این لایه بسته‌ها جمع‌آوری می‌شوند و سپس بازرسی داده از راه رسیده برای یک برنامه کاربردی خاص، می‌تواند با انجام بازرسی عمیق جهت فیلترینگ محتوا، بر عهده گرفته شود.

جدول ۲-۱: مدل OSI [۳]

<p>▪ کاربر را قادر می‌سازد تا دسترسی به شبکه را بپذیرد. رابط کاربر را فراهم می‌کند و خدماتی مانند ایمیل، WWW، FTP و غیره را فراهم می‌کند. Gateway در این لایه قرار دارد. آدرس‌دهی‌های مختلفی بر پایه این خدمات وجود دارد مانند آدرس‌های ایمیل، آدرس‌های UML، آدرس‌های اتصال FTP و غیره.</p>	<p>لایه کاربرد لایه ۷</p>
<p>▪ نحو و معنا اطلاعات تبادل شده میان دو کاربرد را به کار می‌برند. ترجمه نحو میان سیستم‌های مختلف، رمزگذاری، رمزگشایی و فشرده‌سازی را فراهم می‌کند.</p>	<p>لایه ارائه لایه ۶</p>
<p>▪ به عنوان یک کنترل‌کننده شبکه محاوره‌ای عمل می‌کند. برقراری، نگهداری و هماهنگ‌سازی تعامل میان دو سیستم را انجام می‌دهد. کنترل محاوره و خدمات جداسازی محاوره را فراهم می‌کند.</p>	<p>لایه نشست لایه ۵</p>
<p>▪ مسئول پردازش فرآیند تحویل است. QOS را از طریق انتقال اتصال‌گرا و بدون اتصال بسته‌های داده IP در بسته‌های لایه انتقال، فراهم می‌کند. از پروتکل‌هایی مثل TCP، UDP، SPX استفاده می‌کند. آدرس‌دهی پورت، تقسیم‌بندی، نظارت بر اتصال جریان کنترل و نظارت بر خطا را انجام می‌دهد.</p>	<p>لایه انتقال لایه ۴</p>
<p>▪ مسئول انتقال نقطه به نقطه بسته‌های داده در سراسر شبکه اینترنت شامل آدرس‌دهی منطقی و مسیریابی است. از پروتکل‌هایی مثل IP، IPX، NetBEUI استفاده می‌کند. مسیریاب‌ها در این لایه قرار دارند. واحد پیام‌رسانی، بسته‌های داده شامل آدرس‌های IP و شناسایی کامپیوتر در شبکه اینترنت می‌باشد.</p>	<p>لایه شبکه لایه ۳</p>
<p>▪ مسئول انتقال بسته بین دو سیستم در شبکه LAN است. فریم‌سازی، کنترل جریان، کنترل خطا، کنترل دسترسی و آدرس‌دهی فیزیکی را انجام می‌دهد. واحد پیام‌رسانی فریم و آدرس‌دهی از طریق آدرس‌دهی سخت افزار (NIC) است. پروتکل‌هایی مانند اترنت، حلقه نشانه و غیره استفاده می‌شود. پل‌ها و سوئیچ‌ها در این لایه کار می‌کنند.</p>	<p>لایه لینک داده لایه ۲</p>
<p>▪ مسئول انتقال و دریافت بیت‌های داده از/به رسانه، بازنمایی بیت‌ها، نرخ داده و هماهنگ‌سازی است.</p>	<p>لایه فیزیکی لایه ۱</p>