

how many nurses are needed to reduce the average waiting time of a patient in the emergency room to less than 10 min.

fix Goal seeking for savings goal in Rufe

Scenario: want to save 50,000

Known info: you can save 4000 each month

Goal: how many months it takes to reach

goal seeking process

see the total saving goal to 50,000

Result

After applying goal seeking, it will take 13 months to save 50,000

Decision Making with Spreadsheets

Unit-IV

Security First Insurance Keeps its connection with policyholders

* challenge
* solution
* results

* Security First insurance is one of the largest homeowners insurance companies in Florida.

* Headquartered in Ormond Beach. It employs more than 80 insurance professionals to serve its nearly 190,000 customers

Challenges

* Florida's Hurricane exposure
* Security First's commitment
* surge in claims after hurricanes
* Evolving communication channels
* Need for proactive & integrated approach

① Florida's Hurricane exposure

Florida faces the highest exposure to hurricanes in the U.S. with an average of 12 named storms & 9 hurricanes annually impacting property & people.

② Security First's commitment

The company is financially strong enough to withstand multiple natural disasters & promises to support customers "storm after storm, year after year."

③ Surge in claims after hurricanes:
while security first processes too slowly, this number can rise to tens of thousands after a hurricane, creating challenges in managing the influx.

④ Evolving communication channels:
customers now contact the company through various means including phone, email, & social media platforms like Facebook & Twitter, necessitating a more responsive approach.

⑤ Need for proactive & integrated approach:
concerned about response times, security first recognized the need to integrate social media responses into the claims process & document them for regulatory compliance.

Solution

- providing responsive service no matter how customers get in touch
- collaboration with IBM Business partner
- SMC4 solution
- Content Analytics & claims integration
- centralized communication repository
- seamless integration with company's system

* prioritizing communications with access to smarter content
- emergency support
- prioritizing requests
- Smarter Content Analysis
- prioritization & routing
- faster, personalized responses.

1. Providing Responsive Service No Matter How Customers Get in Touch.

* collaboration with IBM Business partner
Security first partnered with a leading IBM Business partner to enhance customer experience by utilizing social media.

* SMC4 solution

The IBM Business partner implemented a solution called Social Media Capture (SMC4) built on "IBM Enterprise Content Management Software", providing four essential:

* capabilities: capture, control, compliance & communication

* the SMC4 solution logs all social networking interaction for security first, capture content, monitors incoming & outgoing messages & archives all communication for compliance review.

- content collector automatically attaches content & attachment
 - For email content back to the policy holder captures email & sends an email receipt.
 - acknowledging receipt.
 - acknowledges & claims. integrates content analytics with enterprise content analysis with enterprise posts & email.
 - Search & analyze customer information & integrates relevant information & integrates relevant documents & it directly into claims process.
 - initiate the claims process.
- centralized communication repository
 - integrating communications from the company's website, emails & the internet.
 - Company's collected onto a "central" file.
 - File Net content are collected onto maintain, control Manager repository to maintain, control link to the appropriate workflows.
- seamless integration with company system
 - integrates early with security solution.
 - The solution integrates with existing applications, databases & processes, enhancing overall claims management efficiency.

- ## 2. Prioritizing communications with Access to Security content
- * Emergency Support: After a hurricane, people whose homes are damaged or destroyed are often forced to leave home quickly with little more than the clothes on their backs. They rely on their insurance companies to promptly provide the necessary support.
 - * Prioritizing Requests: When tens of thousands of policyholders need assistance in a short time, security first must quickly sort & prioritize requests. The content analysis with enterprise search software - helps identify the most critical cases.
 - * Smarter content analysis: The software uses text mining, text analytics, natural language processing & sentiment analysis to analyze emails, social media posts, tweets & comments, detecting words & tones that signal significant distress.
 - * Prioritization & Routing: This analysis enables security first to prioritize messages and direct them to the appropriate person to offer reassurance, address complaints or process claims.

* Faster, Personalized Responses: with an AI system that can respond to smarter content, Security First can respond more quickly, efficiently, & personally to customers, ensuring that those industry customers receive the appropriate level of assistance.

Result:

- * Meeting Regulatory Requirements
- * Controlled Employee Responses
- * Automated Verification & Message Review
- * Tracking & Maintaining Control
- * Expanded Use of Social Media

Meeting Regulatory Requirements: Security

First uses IBM software text analysis capabilities to filter inappropriate incoming communications & monitor outgoing messages, ensuring they meet industry regulation.

Controlled Employee Responses: The system allows Security First to designate specific employees or roles to create & submit responses, ensuring the responses follow company policies & industry standards.

Automated Verification & Message Review: The system automatically verifies the designated personnel & analyzes outgoing message content, flagging ineffective or questionable communications for further review.

Tracking & Maintaining Control: All communication interactions are recorded, enabling Security First to track & manage the process, including controlling which employees respond, their authority level & message content.

Expanded Use of Social Media: With regulatory concerns addressed, Security First confidently expands its use of social media, allowing the company to directly engage with customers & enhance communication opportunities.

Web Mining Overview

The web, as the world's largest data repository, contains vast unstructured textual data in the form of web pages coded in HTML or XML, along with hyperlink connections & usage logs.

Analyzing this data optimizes web usage by personalizing experiences, such as recommending products based on past behavior, & improving content relevance & engagement.

Challenges in Web Mining: Overview

The complexities of the Web

- * Because of its sheer size & complexity mining the web is not an easy undertaking by any means.
- * The web also poses great challenges to effective & efficient knowledge discovery.
- 1. The web is too big for effective data mining.
- 2. The web is too complex.
- 3. The web is too dynamic.
- 4. The web is not specific to a domain.
- 5. The web has everything.

The web is too big for effective data mining.

* The web is so large & growing so rapidly that it is difficult to even quantify its size.

* Because of the sheer size of the web, it is not feasible to set up a data warehouse to replicate, store, & integrate all of the data on the web, making data collection & integration a challenge.

The web is too complex.

* The complexity of a web page is far greater than that of a page in a traditional text document collection. Web pages lack a unified structure.

* They contain far more authoring style & content-variation than any set of books, articles, or other traditional text-based.

The web is too dynamic.

* The web is a highly dynamic information source. Not only does the web grow rapidly but also its content is constantly being updated.

Blogs, news stories, stock market results, weather reports, sports scores, price, company advertisements & numerous other types of information are updated regularly on the web.

The web is not specific to a domain

* The web serves a broad diversity of communities & connects billions of workstations.

Web users have very different backgrounds, interests and usage purposes. Most users may not have good knowledge of the structure of information network and may not be aware of the heavy cost of a particular search that they perform.

The web has everything.

Only a small portion of the information on the web is truly relevant (or useful) to someone. It is said that 99% of the information on the web is useless to 99% of web users.

Although this may not seem, it is true that a particular person is generally interested in only a tiny portion of the web, whereas the rest of the web contains information that is uninteresting to the user & may swamp desired results.

Finding the portion of the web that is truly relevant to a person & the task being performed is a prominent issue in web related research.

* The challenges of web data discovery led to research aimed at improving efficiency.

* While index-based search engines help locate documents using specific keyword, they often return irrelevant (or) too many results & may miss highly relevant documents that don't match the exact search terms.

* Web mining offers a more effective

Web content & Web structure mining

- * web content mining
- * web crawlers
- * illustrative example
- * Central (or) authoritative page & hub web structure mining

Web content mining

- * It refers to the extraction of useful information from web pages.
- * The documents may be extracted in some machine-readable format so that automated techniques can extract some information from these web pages.
- * A company wants to understand customer sentiment about its product.
- * Web crawlers are used to read through the content of a web site automatically. The information gathered may include document characteristics similar to what is used in text mining.
- * Such an automated process of collecting & mining of web content can be used for competitive intelligence.
- * It can be used for information / option collection & summarization, sentiment analysis, & automated data collection & structuring for predictive modeling.

Illustrative example

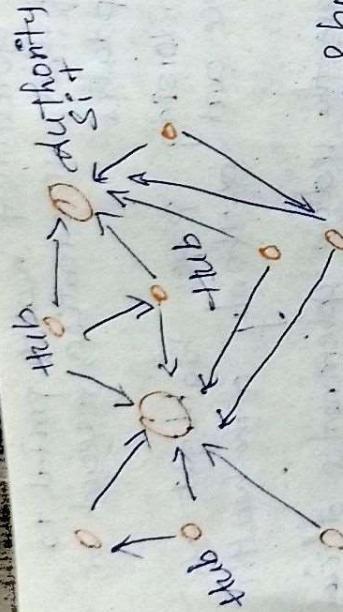
- * Drs. Sharda & Peters developed models to predict Hollywood movie of financial success before release, using data from various websites with different page structures.

- Manually collecting data on thousands of pages was time-consuming & costly of money used. Web content management is poor, so they used spiders to automate the process & spiders to automatically collect, verify & store data in a relational database, ensuring data quality while saving valuable time.
- The automated system collects, verifies & stores data in a relational database, ensuring data quality while saving valuable time.
- Web pages have hyperlinks that connect to other pages. In this case, several which pages are these central pages, worth more important or central pages, worth more links are often considered more authoritative.
- Authoritative web developer includes a link to another page, it acts as an endorsement of the page. It collects links indicating the importance of the page.
- The vast web linkage information offers valuable insights into a relevant, quality & structure of web content, making it a rich resource for web mining.
- Example of authoritative page:
- Websites like Indian Institute of Technology (IITs), India Council of Medical Research (ICMR), UNICEF India might

- link to the government health portal to provide other information, research or news updates.
- These enclosing pages highlight the relevance and authority of the health portal.
- In this case, the Indian government's official health portal is considered the central or authoritative page, as it is endorsed by authoritative source, & is likely to be trusted, reputable source.
- Rank higher in search results.
- The structure of web hyperlink has led to another important category of web pages called a hub.
- It is a hub for one or more web pages that provide a collection of links to authority pages.
- It is a hub because each provider links to authoritative sources, such as academic to authoritative sources, such as academic papers, books & government.
- Studies & data from institutions like NASA.

- as Google & Yahoo!
- * In Search engines like Google & Yahoo
- * Links to a website help determine its authority, meaning the more links pointing to a page, the more popular or trusted it is.

- * In Google's Page Rank algorithm, words like 'authoritative' pages & hubs, we can find high-quality content on specific topics.
- * As authoritative pages link to collections of information, & hubs
- * These authoritative pages allow for improved search results & more accurate, relevant content retrieval from the web.
- * the most popular publicly known & referenced algorithm used to calculate differences between hyperlinks induced hubs & authorities is hyperlink-induced topic search (HITS)



- * This relationship allows for improved search results & more accurate, relevant content retrieval from the web.
- * the most popular publicly known & referenced algorithm used to calculate differences between hyperlinks induced hubs & authorities is hyperlink-induced topic search (HITS)

Web Structure Mining

- * This is the process of extracting information from the links embedded in web documents
- * It is used to identify authoritative pages and hubs, which are the cornerstones of the contemporary pagerank algorithm & one central to popular search engines such

- * In Google's Page Rank algorithm, words like 'authoritative' pages & hubs, we can find high-quality content on specific topics.
- * As authoritative pages link to collections of information, & hubs
- * These authoritative pages allow for improved search results & more accurate, relevant content retrieval from the web.
- * the most popular publicly known & referenced algorithm used to calculate differences between hyperlinks induced hubs & authorities is hyperlink-induced topic search (HITS)

Search Engines

- * It is a software program that searches for documents based on the keywords users have provided.

* The overall goal of a search engine is to return one or more documents/pages that best match the user's query

Anatomy of a Search Engine

* At the highest level, a search engine is composed of two main cycles: a development cycle & responding cycle. While one is interfacing with the web, the other is interfacing with the user.

* One can think of the development cycle as a product production process.

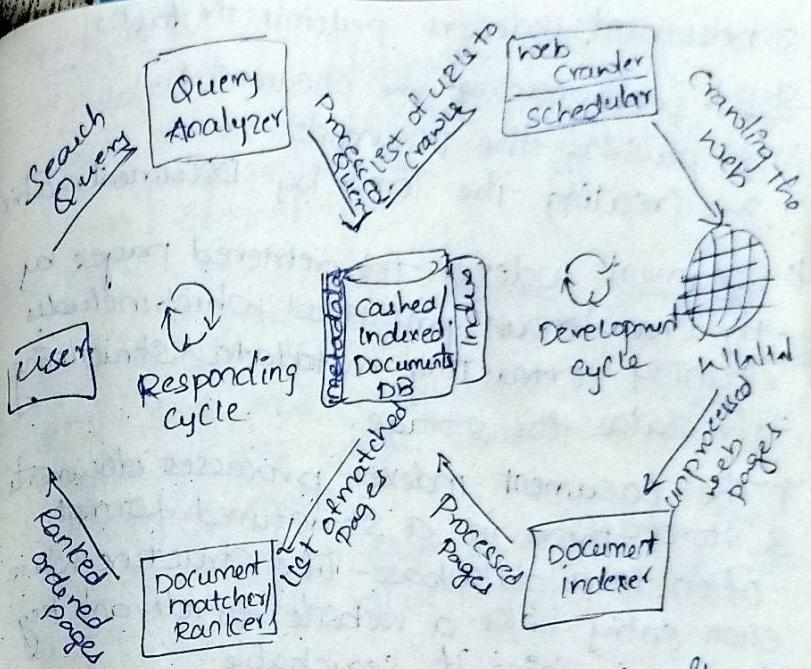
Development Cycle

* It is the interfacing with the worldwide web.

* The two main components of the development cycle are the web crawler & document indexer.

* The purpose of this cycle is to create a huge database of documents/pages organized & indexed based on their content & information value.

* Web crawler: It is the name for the purpose of finding & fetching web pages. Often web crawlers copy all the pages they visit for later processing by other functions of a search engine.



① A web crawler is a piece of software that systematically browses the WWW for the purpose of finding & fetching web pages. Often web crawlers copy all the pages they visit for later processing by other functions of a search engine.

② As the crawler visits these URLs it... identifies all the hyperlinks in the page & adds them to the already existing list of URLs to visit.

③ Document Indexer: The retrieved pages by the crawler are processed, which includes cleaning, extracting metadata & structuring the data for storage.

Performs the task

- 2. Document indexers & the documents
 - 1: pre processing the documents
 - 1: crawling the web pages
 - 2: parsing the html
 - 3: creating the term-document matrix
 - 2: creating the inverted index
 - 1: document indexers: the retrieved pages are crawled, processed, which include cleaning, extracting, structuring, normalizing, and indexing the data for storage.
 - 2: document indexers process the document in a structured format, & stores them in a data base-like structure, often in a data base to relevant to each entry links a website to relevant words, making it searchable.
 - 3: document indexers
 - 1: document indexers, database, metadata
 - 2: stores preprocessed pages by structure, ensuring up-to-date document indexers, information for the search, & search cycle
 - 3: responding to user queries
 - 3. Document retrieval
 - 1: stores preprocessed pages with relevant page & context type for quick retrieval
 - 2: document indexers
 - 1: stores preprocessed pages, that have a repository of web pages, that have already been crawled & indexed. contains structured information such as page with page records.

Metadata additional information associated with the document including:

- metainformation or additional information about a document, including:
each document, including
→ page, relevance scores
→ content type

3. Creating indexer: the retrieved page is document processed, which includes crawling, extracting, metadata & structure, cleaning & storage.
4. Document indexer processes document in a structured format, stored in a data base-like structure and often in a website to relevant by each entry linking it to the document.

Words, make metadata indexed. Documents, Database,
stores the structure, processes pages by
indexing, ensuring up-to-date
information for the
searchable, information for the
responding, cycle

- * It stores preprocessed & indexed web page with metadata like relevant content & context type for quick retrieval
- * Indexed documents DB
- * A repository of web pages that have already been crawled & indexed.
- * Structured information such as page titles, records.

Response cycle with the user

- It is the interfacing of the responding to the two main components of the responding in the query analyzer & document cycle are the query matcher / ranker
- user & search Query: users enter their search query, reflecting their information search.
- Query Analyzer: processes the query by extracting key words & connecting spelling, applying NLP for better understanding & applying NLP for better understanding of the query
- Keyword extraction: identifies key terms & phrases from the query
- NLP Techniques: understands the query, information, transactional, non-transactional, intent, relevance, or navigational
- Spelling corrections: Fixes misspelled words
- Stemming: Reduces words to their root form. for ex: "running", "run", & "runner" are reduced to "run" to unify similar queries.

stopword Removal : eliminates common words like "the", "is", "and" that don't affect search intent

Methods : SEO involves optimizing website content, HTML & coding to match relevant keywords & remove barriers for search indexing.

Processed Query is sent to the indexed refined database for matching the documents that are already crawled & stored here for quickly retrieving indexed queries during search.

Matched Ranks : computes the query document scores & assigns relevance against indexed keywords, authority & auto scores based on their content & information.

Retrieved Pages : Returns ranked documents as search results, completing the cycle as part of this development cycle.

The purpose of this development cycle is to create a huge database of documents organized & indexed based on their content & information value.

Search Engine Optimization

Purpose : SEO aims to improve the visibility of a website in unpaid search engine results, increasing traffic by ranking higher in search results.

- * Backlinks are links from other websites that point to your site. To improve search engine rankings, websites aim to get other, trusted sites to link back to their content, which signals to search engines that the site is valuable & relevant.
- * Search engine crawling : crawlers automatically visit websites, gather content like text, images, links & metadata & index it.
- * Ranking Tactics : SEO tactics include using relevant keywords, updating content regularly & optimizing metadata to boost search rankings & drive traffic.

SEO techniques can be broadly divided into two categories

1. white-hat SEO
2. black-hat SEO

White-hat SEO

- These techniques align with search engine guidelines & involve no deception
- white-hat SEO focuses on creating quality content for users, not just for search engines & making it easily accessible to crawlers
- It is about ensuring that the content indexed by search engines matches what users will see
- white-hat methods generally lead to long-term sustainable results.

Black-hat SEO

- These methods aim to improve rankings using deceptive techniques that are not approved by search engines
- ex. include hiding text in the same color as the background (css) using cloaking
- Black-hat techniques can result in penalties from search engines, which include ranking reductions or removal from the search engines' index

- * The webpage has visible content for users like "explore our premium collection of coffee makers designed to suit every taste!"
- * when a search engine crawler visits the website, it is served a page filled with keyword-rich content & detailed description.

- * The website may temporarily rank higher using cloaking & keyword manipulation but incurs severe penalties like demotion or removal from search results. If detected by search engines.

Web Usage Mining (Web Analytics)

- * web usage mining is the process of extraction of useful information from data generated through web page visits & transactions
- * Analysis of the information collected by web servers can help us better understand user behavior
- * Analysis of this data is often called clickstream analysis. By using the data & text mining techniques, a company might be able to identify interesting patterns from the clickstreams.
- * ex. 60% of visitors who searched for "hotels in delhi" had searched earlier for "flights to delhi". Such information could be useful in determining where to place online advertisement
- * clickstream analysis might also be useful for knowing when visitors access a site
- * The process of extracting knowledge from clickstream data & how to generate knowledge is used to improve the process, website & most important, increase the customer value

Web Analytics Technologies

Web Analytics tools are increasingly popular for measuring, collecting & analyzing internet data to optimize web usage & transform online business.

Web analytics supports e-business & market research by tracking traffic changes, visitor numbers, page views & trends to evaluate e-commerce effectiveness.

After launching a discount campaign, web analytics can track a spike in website visits & purchases helping a retail business evaluate the campaign's success.

There are two main categories of web analytics

- Off-site & on-site
- Off-site web analytics refers to web measurement & analysis about you & your products that takes place outside your web site
- Includes the measurement of a website's potential audience, share of voice & buzz that is happening on the internet

- On-site web-analytics there are two technical ways of collecting the data
 - The first & more traditional method is the server log file analysis - where the web-server records file requests made by browsers.



* The five points aims to analyze user behavior & enhance marketing strategies through Segmentation & targeted marketing

1. Keywords
2. Geography
3. Time of Day
4. Landing page profiles

Conversion statistics

Each organization will define a "conversion" according to its specific marketing objectives. Some web analytic programs use the term "goal" to benchmark certain web objectives.

1. New visitors
2. Returning visitors
3. Leads
4. Sales/Conversions
5. Abandonment/exit rates

Web Analytics Maturity Model

* The web analytics maturity model is a framework that helps organizations evaluate their level of expertise in using web analytics.

* It outlines different stages, from basic tracking of website data to advanced data-driven decision-making.

A maturity model is formal depiction of critical dimensions & their competency levels define the maturity level of an organization in that area of practice

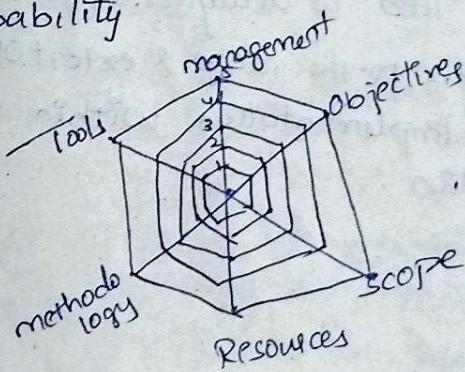
* ex: a small online retailer initially relies on guess work to decide which products to stock.

* As a business matures, it implements a data-driven approach by analyzing customer purchasing trends, setting up automated inventory management systems & optimizing stock levels based on predictive analytics.

* The simple business analytics maturity model, moving from simple descriptive measures to predicting future outcomes, to obtaining sophisticated decision systems

* not perhaps the most comprehensive model was proposed by Stephane Hamel

* Hamel used six dimensions integration & for each dimension he used six levels of capability



Dimension

① Management

2. Scoping

1. Analytic Team & expertise
2. Methodology
3. Tools
4. Six levels are indications of analytical maturity ranging from "0-Analytic impaired" to "5-Analytical competitor"

0. Impaired
1. Initiated
2. Operational
3. Integrated
4. Competitor
5. Addicted

Management: It refers to the leadership's involvement in web analytics initiatives & their ability to create a data-driven culture & objectives. It refers to the clarity of goals & objectives tied to analytics efforts.

Scope: refers to the range & extent of web analytics implementation within an organization.

② Collected (0): Organizations out-there have little/no web analytics or goals, no web analytics are reported (0) inaccurate data collection.

Initiated (1): Value of analytics tools are collected. Core data is collected, but there is no analysis.

Operational (2): Web analytics are regular & used with a focus on tracking user behavior, user reporting & some insights are used to improve decision making.

Integrated (3): Web analytics are integrated across multiple channels & data is used holistically to optimize business strategies & customer experiences.

Competitor (4): The organization uses analytics to benchmark performance against competitor.

To benchmark performance against competitor.

Competitive edge (5): Analytics are deeply embedded in every aspect of the business, with constant optimization, real-time data use.

Constant optimization, real-time data use.

Web Analytics Tools

There are plenty of web analytics applications on the market. The most popular free-web analytics tools.

① Google Web Analytics (GOOGLE.COM/ANALYTICS)

This is a service offered by Google that generates detailed statistics about a website's traffic & traffic sources & measures conversion & sales.

② YAHOO! WEB ANALYTICS (YAHOO!ANALYTICS.YAHOO.COM)

Yahoo! web analytics is offered to the dominant Google Analytics. It is a comprehensive web analytics tool, it has graphs, custom-designed reports & time data tracking.

③ OPEN WEB ANALYTICS (OPENWEBANALYTICS.COM)

OWA is a popular open source web analytics software that anyone can use to track & analyze how people use websites & applications.

④ FIRESTAT (FIRESTATS.COM)

Firestats is a simple & straightforward web analytics application written in PHP/MySQL. It supports numerous platforms & set-ups including C# sites, Django sites.

⑤ SITEMETER (SITEMETER.COM)

Sitemeter is a service that provides count & tracking information for websites. It uses JavaScript or HTML to track visitors information. Site meter provides website owners with information about their visitors, including how they reached the site & the date & time of their visit & more.

⑥ ANSTATS (ANSTATS.Org)

It is an open source web analytics reporting tool, suitable for analyzing data from internet services such as web, streaming media, mail & FTP service. It parses & analyzes server log files, producing HTML reports.