গুগল সার্চের সহজ কৌশল

গুগল সার্চের সহজ কৌশলের শেষ পর্বে আলোচনা করা হয়েছে কীভাবে গুগল সার্চ কাজ করে। এতদিন গুগল সার্চের বিভিন্ন কৌশল নিয়ে আলোচনা করা হয়েছে, অথচ কিভাবে গুগল সার্চ কাজ তাই যদি জানা না থাকে তাহলে কেমন হয়?

হাসান মাহমুদ

গুগল সার্চ যেভাবে কাজ করে

গুগল কিভাবে কাজ করে সেটা এখনও সবার মধ্যে বিস্ময়কর। এতগুলো ওয়েবসাইটের মধ্য থেকে গুগল কিভাবে তালিকাবদ্ধ করে এদেরকে সামনে নিয়ে আসে। গুগলের জন্ম দুটি ব্যক্তির হাত ধরে, যাদের একজন হচ্ছেন ল্যারি পেইজ এবং আরেকজন সের্গেই ব্রিন।

গুগলের মূলমন্ত্র হলো 'বিশ্বের তথ্য সন্নিবেশিত করে তাকে সবার জন্য সহজলভ্য করে দেয়া', যেখানে গুগলের অপ্রাতিষ্ঠানিক মূলমন্ত্র হলো 'Don't be evil'। গুগল সারা বিশ্বে বিভিন্ন ডাটা সেন্টারে প্রায় এক মিলিয়ন সার্ভার চালায় ও প্রতিদিন এক বিলিয়নের ওপর সার্চের অনুরোধ এবং প্রায় ২৪ পেটাবাইট ব্যবহারকারীর মাধ্যমে তৈরি করা ডাটা প্রক্রিয়াজাত করে। ২০০৯ সালের সেপ্টেম্বর অনুযায়ী এলেক্সা আমেরিকার সবচেয়ে বেশি ব্যবহার করা ওয়েবসাইটের তালিকায় স্থান দেয় গুগলকে।

আমরা বর্তমানে গুগলের যে হোম পেজটি দেখতে পাই সেটি আপডেট করা হয়েছিল ২০১১ সালের ১৪ ফেব্রুয়ারি। আমরা সবাই প্রতিদিনে কয়েকবার গুগলে সার্চ দিয়ে থাকি এবং নিজেদের পছন্দমতো সমাধান পাই। এবার জেনে নেই গুগলের সার্চ কীভাবে কাজ করে।

০১. গুগলবট নামে গুগলের এক ধরনের ওয়েব ক্রলার আছে, যার কাজ হচ্ছে ওয়ার্ল্ড ওয়াইড ওয়েবের মাধ্যমে বিশ্বের সব ধরনের ওয়েবপেজ ভিজিট করা। যদি আপনার সাইটের সার্চিং বন্ধ করে রাখেন, তাহলে গুগল সার্চ ইঞ্জিন আপনার সাইট সার্চ করতে পারবে না।

০২. এ বটগুলো সবসময় আপনার পছন্দের সমাধান দিতে ব্যস্ত থাকে। তাই তারা প্রতিনিয়ত প্রত্যেকটি ওয়েবসাইটের পেজগুলো ক্রলিং করে, এমনকি একই পেজ আবার ক্রলিং করে, যাতে কোনো তথ্য বাদ না যায়।

০৩. গুগলবট সবচেয়ে বেশি সে সাইটটি ক্রল করে, যা প্রতিনিয়ত পরিবর্তিত হয়। যেমন একটি কোম্পানির সাধারণ ওয়েবসাইট থেকে একটি নিউজ সাইট সবচেয়ে বেশি পরিমাণে ক্রলিং হয় (কারণ নিউজ সাইটগুলো প্রতিনিয়ত আপডেট হয়ে থাকে)।

০৪. গুগলবট একটি সাইটের প্রত্যেকটি লিঙ্ক শনাক্ত করে রাখে। এ লিঙ্কগুলোকে গুগলবট তাদের একটি লাইনে বিন্যস্ত করে রাখে এবং পরে কোনো এক সময় এ লিঙ্কগুলো ভিজিট করে। তাই ওয়েবসাইটে বেশি লিঙ্ক ব্যবহার করলে তা পেজের র‌্যাঙ্ক বাড়াতে ভূমিকা রাখে।

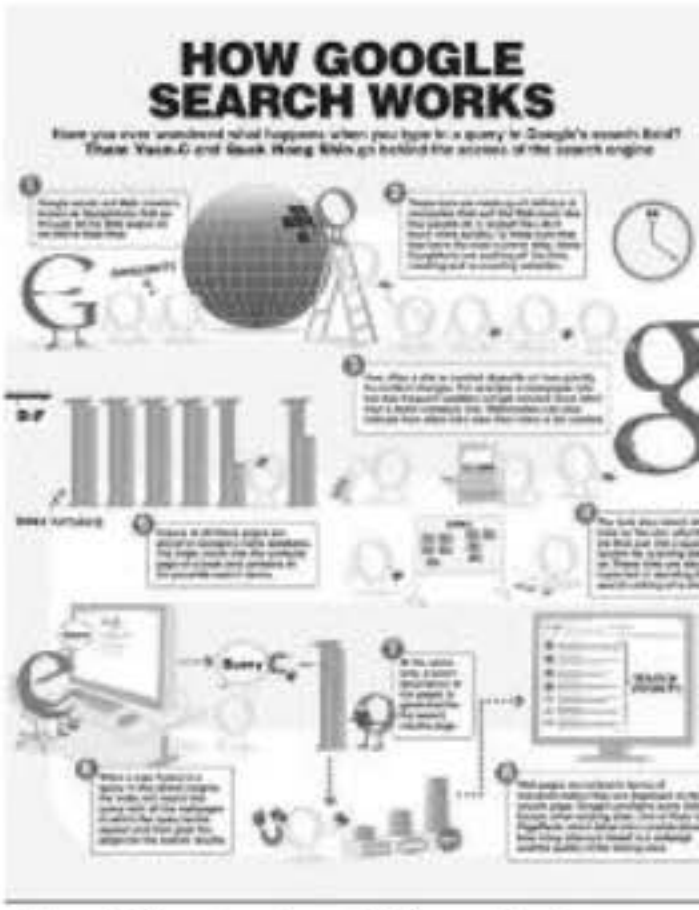
০৫. গুগলবট তাদের ভিজিট করা প্রত্যেকটি পেজকে একটি সূচিপত্রে সাজিয়ে রাখে। এ সূচিপত্রটি আমাদের বইয়ের সূচিপত্রের মতো। যেমন অধ্যায় এক : বাংলাদেশের ইতিহাস-পেজ নম্বর : ১, ঠিক এমনভাবেই গুগলবট প্রত্যেকটি পেজকে একেকটি ক্যাটাগরির সূচিপত্রে সাজিয়ে রাখে।

০৬. যখনই আপনি কোনো কিছু অনুসন্ধান করেন, সেই অনুসন্ধান অনুযায়ী সূচিপত্র থেকে তথ্য নিয়ে সার্চের সমাধানে প্রকাশ করে। যেমন আপনি কোনো দোকানে গিয়ে বললেন, ভাই আমার টুথপেস্ট লাগবে। দোকানদার ঠিক তখনই টুথপেস্টের জন্য সাজানো তাক থেকে অনেকগুলো টুথপেস্ট নিয়ে আপনার সামনে রাখবে। এবার পছন্দ করার দায়িত্ব আপনার।

ঠিক তেমনি ধরুন গুগলে সার্চ দিলেন 'বাংলাদেশী টেক' শব্দটি দিয়ে। এবার গুগল তার সূচিপত্রে রাখা 'বাংলাদেশী টেক' ক্যাটাগরি থেকে আপনাকে সব তথ্য খুঁজে দেবে। এবার আপনি পছন্দ মতো সাইটে ভিজিট করতে পারবেন।

০৭. ঠিক একই সময় গুগল সার্চ ইঞ্জিন প্রত্যেকটি সাইটের সংক্ষিপ্ত বর্ণনা দিয়ে রাখে, যা সার্চিংয়ের সময় দেখানো হয়। আপনি খেয়াল করে দেখবেন, কোনো কিছু সার্চ দিলে অনেকগুলো সাইটের নামের নিচে সংক্ষিপ্ত বর্ণনা দেয়া থাকে।

০৮. গুগল তার একই ক্যাটাগরির সূচিপত্রে একই ধরনের ওয়েবসাইটকে রাখার ক্ষেত্রে অনেকগুলো বিষয় লক্ষ করে। গুগলের এ ওয়েবসাইট র‌্যাঙ্কিং কিসের ভিত্তিতে হয়, তা সবার কাছে অজানা। তবে বলা হয়ে থাকে ২০০টি ব্যাপারের দিকে গুরুত্ব দিয়ে গুগল এ র‌্যাঙ্কিং করে। এর মধ্যে একটি হচ্ছে গুগল পেজ র‌্যাঙ্ক। কতগুলো ওয়েবসাইট একটি পেজের মধ্যে লিঙ্ক করা আছে এবং সেই লিঙ্ক করা সাইটগুলোর গুণ কতটা ভালো, তার ওপর ভিত্তি করে পেজ র‌্যাঙ্ক করা হয়।



ইনফোগ্রাফিটি দেখলে গুগলের সার্চ ইঞ্জিনের পদ্ধতি খুব সহজেই বোঝা যায়

বর্তমানে গুগল কয়েক লাখ সার্ভার ব্যবহার করে। গুগলের কৌশল হচ্ছে কাস্টমাইজ করা অপারেটিং সিস্টেমযুক্ত কম দামী সিস্টেম ব্যবহার করা। অপারেটিং সিস্টেমটি লিনআক্স। সার্ভারগুলো ডকুমেন্ট সার্ভার, অ্যাড সার্ভার ইত্যাদি বিভিন্ন ভাগে বিভক্ত। সার্ভারগুলোতে ডাটা ৬৪ মেগাবাইট ব্লকে স্টোর করা থাকে। ডাটার নিরাপত্তা নিশ্চিত করতে প্রতিটি ডাটা তিনটি করে কপি করা হয় এবং স্টোর করা হয় আলাদা পাওয়ার সাপ্লাই সংবলিত মেশিনে। একটি পাওয়ার সাপ্লাই লাইনে পরিচালিত কোনো সার্ভারে একই ডাটার দুটি কপি থাকে না এবং ডাটাগুলো এমনভাবে বণ্টন করা হয়, যাতে যেকোনো দুটি সার্ভারে কখনই একই ধরনের ডাটা থাকে না। মানে বিষয়টি এরকম-কোনো সার্ভারে যদি ডি-মেইল, ইনডেক্স ও আর্থের ডাটা থাকে। অন্যকোনো সার্ভারে ঠিক এই তিন ধরনের ডাটা থাকবে না। হয় জি-মেইল, ইনডেক্স, ইমেজ অথবা ইনডেক্স, আর্থ, গুগল ডক এরকম। দ্বিতীয় কোনো সার্ভার পাবেন না যেটাতে জি-মেইল, ইনডেক্স ও আর্থের ডাটা আছে।

গুগল প্রধানত তিনটি ভাগে বিভক্ত। ০১. গুগলবট, ০২. ইনডেক্সার ও ০৩. কোয়েরি প্রসেসর।

গুগলবট : গুগলবট ওয়েব থেকে পেজ ►

সংগ্রহ করে। এর কার্যপদ্ধতি অনেকটা আমাদের ব্যবহার করা ওয়েব ব্রাউজারের মতোই। গুগলবটও ওয়েব সার্ভারে ব্রাউজারের মতো পেজ রিকোয়েস্ট পাঠায়। সার্ভার থেকে পেজগুলো পাঠানো হলে সেগুলো স্টোর করে। আমাদের ব্রাউজারের মতো হলেও গুগলবট অনেক বেশি দ্রুতগতিসম্পন্ন। অসংখ্য কমপিউটারের সমন্বয়ে গুগলবট একসাথে কয়েক হাজার পেজ রিকোয়েস্ট পাঠাতে পারে। অনেক দুর্বল সার্ভার গুগলবটের এ বিপুল সংখ্যক রিকোয়েস্ট রেসপন্ড করার সাথে সাথে সাধারণ ইউজারদের রিকোয়েস্ট রেসপন্ড করতে পারে না। তাই সাধারণ ব্যবহারকারীদের প্রাধান্য দিয়ে গুগলবটকে পূর্ণ ক্ষমতায় চালানো হয় না।

গুগলবট নতুন ইউআরএল সংগ্রহ করে প্রধানত দুটি উপায়ে। ০১. http://www.google.com/ addurl.html-এ পাতায় সাবমিট করা পেজ। ০২. ওয়েব ক্রলিংয়ের মাধ্যমে।

গুগলবট যখন একটি পেজ সংগ্রহ করে, তখন এ পেজে পাওয়া লিঙ্কগুলো তার ক্রলিং তালিকায় যোগ হয়। এ পদ্ধতিতে একই লিঙ্ক অসংখ্যবার আসে, কিন্তু গুগলবট সেগুলোকে বাদ দিয়ে একটি তালিকা তৈরি করে যাতে সবচেয়ে কম সময়ে পুরো ওয়েবকে কভার করা সম্ভব। এ ব্যবস্থাকে বলে ডিপ ক্রলিং। কোন পেজ কত দ্রুত পরিবর্তন হয় সেটি ঠিক করা গুগলবটের অন্যতম প্রধান দায়িত্ব। গুগল ডাটাবেজকে আপডেট রাখার ক্ষেত্রে এটি সবচেয়ে বেশি জরুরি। গুগলবট কোনো পেজে পরিবর্তনের একটি ফ্রিকোয়েন্সি বের করে এবং সেই হিসেবে ঠিক করা হয় যে গুগলবট কত সময় পরপর কোনো পেজ ক্রলিং করবে। কারণ যে পেজ মাসে একবার পরিবর্তন হয় সেটা কয়েক ঘণ্টা পরপর ক্রলিং করা সময় নষ্ট ছাড়া আর কিছুই নয়। সবসময় পরিবর্তনশীল সাইটগুলো কয়েক ঘণ্টা পরপর ক্রলিং করা হয়। দৈনিক পত্রিকাগুলো প্রতিদিন আর বাংলাদেশের বেশিরভাগ সরকারি সাইটের মতো পেজগুলো মাসে একবার। ডাটাবেজে আপডেট করার এ ক্রলিংকে ফ্রেশ ক্রলিং বলে।

গুগল ইনডেক্সার : গুগল ইনডেক্সারের কাজ তুলনামূলকভাবে সহজ। গুগলবট ইনডেক্সারকে ক্রলিং করা পেজগুলোর ফুল টেক্সট দেয়। ইনডেক্সার সার্চ টার্মগুলোকে বর্ণমালা অনুক্রমে সাজায় এবং কোন টার্ম কোথায় আছে তা সেভ করে রাখে। কিছু পরিবর্তনও আনা হয় পেজগুলোতে। কিছু বিরাম চিহ্ন বাদ দেয়া হয়। একের অধিক স্পেস থাকলে সেটাও বাদ দেয়া হয়। ইংরেজির ক্ষেত্রে বড় হাতের অক্ষরগুলোকে ছোট হাতের অক্ষরে পরিবর্তন করা হয়।

গুগল কোয়েরি প্রসেসর : এটি সর্বশেষ অংশ। এটাই আমাদের সার্চ রেজাল্ট প্রসেসিং করে। কোয়েরি প্রসেসর কয়েকটি অংশে বিভক্ত- ইউজার ইন্টারফেস, কোয়েরি ইঞ্জিন, রেজাল্ট ফরম্যাট ইত্যাদি। গুগলের ওয়েবপেজ র‌্যাঙ্কিং সিস্টেমের নাম পেজর‌্যাঙ্ক। যে পেজের পেজর‌্যাঙ্ক যতবেশি সেটা সার্চ রেজাল্টে তত উপরে থাকে। পেজর‌্যাঙ্ক নির্ধারণ করা হয় অনেক কিছু বিচার করে। পেজটির জনপ্রিয়তা, সার্চ টার্মের অবস্থান ও আকার, অন্য পেজে টার্মটি কতবার আছে, একাধিক টার্ম হলে শব্দগুলোর মাঝে দূরত্ব, পেজটি কতদিন ধরে ওয়েবে আছে ইত্যাদি অনেক কিছু বিচার করে পেজর‌্যাঙ্ক নির্ধারণ করা হয়। একই সাথে গুগল সার্চ টার্মগুলোর পারস্পরিক সামঞ্জস্যতা বিচার করে। এর ভিত্তিতে গুগলের spelling-correction সিস্টেম কাজ করে। গুগলবট যেহেতু টেক্সটের সাথে পেজ কোডও ক্রলিং করে, তাই ইউজার চাইলে সার্চ টার্মটির অবস্থানও নির্দিষ্ট করে দিতে পারেন যে সেটি লিঙ্কে থাকবে, টাইটেলে থাকবে না লেখায় থাকবে। শুধু টার্মের উপস্থিতির ওপর ভিত্তি করে রেজাল্ট না দেয়ার কারণেই গুগলের সার্চ রেজাল্টের মান এত উন্নত।

১৯৯৮ সালে গুগলে ব্যবহার হওয়া যন্ত্রপাতি

* দুটি ডুয়াল পেন্টিয়াম টু প্রসেসর ৩০০ মেগাহার্টজ সার্ভার যাদের ছিল ৫১২ মেগাবাইট র‌্যাম।
* চারটি প্রসেসরযুক্ত ৫১২ মেগাবাইট র‌্যামের একটি এফ৫০ আইবিএম আরএস৬০০০ কমপিউটার।
* একটি ডুয়াল প্রসেসর সান আল্ট্রা২ ৫১২ মেগাবাইট র‌্যামযুক্ত কমপিউটার।
* কয়েকটি হার্ডডিস্ক, প্রতিটি ৪ থেকে ৯ গিগাবাইট। মোট ৩৫০ গিগাবাইট।

বর্তমানে গুগলের রয়েছে বিশ্বের বিভিন্ন স্থানে কয়েক লাখ সার্ভার। মোট ডাটার পরিমাণ প্রায় ৩০০ টেরাবাইট। ২০০৪ সাল থেকে গুগল ইন্টেলের পরিবর্তে এএমডি প্রসেসর ব্যবহার করছে বিদ্যুৎ সাশ্রয়ের জন্য।

এতক্ষণ জানলাম কিভাবে গুগল সার্চ কাজ করে। এবার দেখে নেয়া যাক নতুন অথবা ভবিষ্যৎ গুগলের ধরন। গুগল সার্চ হবে এখন 'কথোপকথন'-এর মাধ্যমে।

গুগল সার্চ কথোপকথন : আপনি ইন্টারনেটে কোনো কিছু সার্চ করছেন অথচ আপনার মনে হচ্ছে আপনি কোনো ব্যক্তির সাথে কথা বলছেন বা কোনো একটি ভয়েস আপনার প্রশ্নের উত্তর দিচ্ছে। ভাবছেন কিভাবে সম্ভব? এ অসম্ভবকে সম্ভব করতে চলেছে বিখ্যাত সার্চ ইঞ্জিন গুগল।

গুগলের ঘোষণা মতে, এখন থেকে সার্চের জন্য একজন ইউজার যেকোনো প্রশ্ন ভয়েসের মাধ্যমে করতে পারবেন এবং উত্তর পাবেন ভয়েসের মাধ্যমে। এক কথায় সার্চটি সম্পূর্ণ হবে একটি আলাপচারিতার মাধ্যমে। গুগল ইতোমধ্যে নতুন এ ভয়েস সার্চ প্রক্রিয়ার সফল পরীক্ষা সম্পূর্ণ করেছে। গুগলের নতুন এ সার্চ পদ্ধতি শুধু কমপিউটারে ইন্টারনেট ব্যবহারকারীদের জন্য, তবে গুগল বলছে আমরা মোবাইল ইউজারদের এ সুবিধা দেয়ার জন্য কাজ করে যাচ্ছি।

আমেরিকার সান ফ্রান্সিসকোতে গুগলের ডেভেলপার কনফারেন্স 'গুগল আই/ও'তে এ ঘোষণা দেন সার্চ বিভাগের ভাইস প্রেসিডেন্ট জননা রাইট। রাইট ওই সময় একটি কনভারসেশনাল সার্চ উপস্থিত সবার মধ্যে করেন। তিনি ভয়েসের মাধ্যমে গুগলের কাছে সান্তা ক্রুজে কী করা হয় দেখতে চান। গুগল তখন ওই এলাকার জনপ্রিয় কার্যক্রমের একটি তালিকা প্রদর্শন করে। রাইট ওই তালিকা থেকে একটি অপশন সিলেক্ট করে প্রশ্ন করেন- এই জায়গা থেকে সান্তা ক্রুজ কত দূরে'। সাথে সাথে গুগল অডিও ভয়েসের মাধ্যমে উত্তর দেয় এবং একই সাথে ব্রাউজারে প্রদর্শন করে সান ফ্রান্সিসকো থেকে সান্তা ক্রুজের দূরত্ব।

গুগলের এ 'কনভারসেশনাল সার্চ' খুব শিগগিরই ব্যবহারকারীদের জন্য উন্মুক্ত করা হবে। সার্চের নতুন এ সুবিধা শুধু গুগলের ব্রাউজার গুগল ক্রোমে ব্যবহার করা যাবে। তবে অবশ্যই ব্যবহারকারীর কমপিউটারে ব্যবহারযোগ্য একটি মাইক্রোফোন থাকতে হবে।

গুগলের মতে, নতুন এ সার্চ সুবিধা একজন ব্যবহারকারীর প্রয়োজনের তুলনায় অনেক বেশি তথ্য দেবে। এ সার্চ সুবিধায় আরও থাকছে পাবলিক ট্রানজিট, সঙ্গীত অ্যালবাম, বই, টিভি শো, রিমাইন্ডার ইত্যাদি। গুগল নতুন এ সার্চ পদ্ধতিকে জ্ঞান চিত্র (Knowledge Graph) বলে আখ্যায়িত করেছে। যার ফলে একজন ব্যবহারকারী প্রতিটি সার্চের বিপরীতে কয়েক বস্তু তথ্য দেখতে পারবেন।

উইকিপিডিয়ার তথ্যমতে, প্রতিদিন গড়ে তিনশ মিলিয়ন ইউজার এ সার্চ ইঞ্জিনে দুই বিলিয়ন সার্চের কাজ করে থাকেন। উল্লেখ্য, ফেসবুক এ বছরের ১ মার্চ থেকে গ্রাফ সার্চ নামে একটি অপশন চালু করে। ফেসবুকের এ গ্রাফ সার্চ অপশন গুগলের কর্তাব্যক্তিরা সহজে মেনে নিতে পারেননি। গুগলের সিইও ল্যারি পেজ এক প্রশ্নের জবাবে বলেন- গুগল ফেসবুকের গ্রাফ সার্চ নিয়ে মোটেও চিন্তিত নয়। পেজ আরও বলেন, সত্যিই তারা (ফেসবুক) তাদের প্রোডাক্টের জন্য একটি খারাপ কাজ করছে। ফেসবুকের বর্তমান গ্রাহক সংখ্যা ১.১১ বিলিয়ন। প্রযুক্তি বিশ্লেষকদের মতে, গুগল তার জনপ্রিয়তা ধরে রাখতেই নতুন এ সার্চ পদ্ধতি।

এখন শুধু অপেক্ষার পালা কোনটি বেশি জনপ্রিয় হয়- ফেসবুকের গ্রাফ সার্চ নাকি গুগলের কনভারসেশনাল সার্চ

ফিডব্যাক : faisalb01@gmail.com