পিপীলিকা : দেশের প্রথম পূর্ণাঙ্গ সার্চ ইঞ্জিন

ইমদাদুল হক

বাংলা ১৪২০ সালের সূর্য ওঠার আগেই পিদিমের আলোয় উদ্ভাসিত হলো ভার্চুয়াল জগত। জাতীয় কবি কাজী নজরুল ইসলামের বিদ্রোহী কবিতার সুরে নতুন প্রজন্ম ঘোষণা করল 'চির উন্নত মম শির'। বর্ষবরণের আগের রাতে বিশ্বের সব বাংলাভাষীকে উপহার দিল দেশের প্রথম পূর্ণাঙ্গ সার্চ ইঞ্জিন 'পিপীলিকা'।

পিপীলিকা, কথ্য ভাষায় যাকে আমরা বলি পিঁপড়া। সামাজিক এই পোকাটি প্রাণীদের মধ্যে অধিকতর বুদ্ধিমান এবং কর্মঠ হিসেবে পরিচিত।

প্রাণিবিজ্ঞানীদের সর্বশেষ গবেষণা অনুযায়ী, পৃথিবীতে রয়েছে ২২ হাজার প্রজাতির পিঁপড়া। প্রযুক্তির ছোঁয়ায় বাংলাভাষায় তথ্য খুঁজে দিতে গত ১৩ এপ্রিল থেকে বাংলাভাষায় যুক্ত হলো পিপীলিকা নামের এক সার্চ ইঞ্জিন। নিজস্ব প্রজাতির মতোই এই পিপীলিকা মজুদ করেছে নানামাত্রিক তথ্য। মাতৃভাষা বাংলার পাশাপাশি বিশ্বভাষা ইংরেজিতেও এই উন্মুক্ত ওয়েবসার্ভিসটি সারাদেশের সাম্প্রতিক গ্রহণসাধ্য তথ্য অনুসন্ধান করতে সহায়তা করে। স্বয়ংক্রিয়ভাবে বিশ্লেষণ ও সংরক্ষণ করে দেশের প্রধান বাংলা ও ইংরেজি পত্রিকার সংবাদ, বাংলা ব্লগ, বাংলা উইকিপিডিয়া ও সরকারি তথ্য। সংযুক্ত করা হয়েছে ভুল বানান লিখেও স্বয়ংক্রিয় শুদ্ধ বানান অনুসন্ধান সুবিধা।

শ্রেণীকক্ষের গবেষণা থেকে বাস্তবে

সার্চ ইঞ্জিনটি নিয়ে শাবিপ্রবি গত কয়েক বছর ধরে কাজ করে আসছে। আর পিপীলিকা সার্চ ইঞ্জিনটি শাহজালাল বিজ্ঞান ও প্রযুক্তি বিশ্ববিদ্যালয়ের একটি থিসিসের ফল। পিপীলিকার আগের ভার্সনটি 'একুশে ফিন্যান্স' নামে পরিচিত ছিল। এটিতেও পিপীলিকার গবেষকেরা কাজ করেছিলেন। একুশে ফিন্যান্স সিলেটের ডিজিটাল ইনোভেশন ফেয়ার ২০১০-এ প্রথম পুরস্কার পেয়েছিল। তবে সেটাতে পিপীলিকার অনেক ফিচার অনুপস্থিত ছিল। বিশেষ করে স্বয়ংক্রিয়ভাবে বাংলা অভিধানের প্রয়োগ ও বিভিন্ন ধরনের কাস্টমাইজড সার্চ একমাত্র পিপীলিকাই দিতে পারে। এ ছাড়া পিপীলিকাতে বাংলার তথ্য ফলাফল করার জন্য সম্পূর্ণ নতুন ধরনের অ্যালগরিদম ব্যবহার করা হয়েছে।

পিপীলিকার অন্দরমহলে

পিপীলিকার উন্মুক্ত ওয়েবসার্ভিসটি সারাদেশের সাম্প্রতিক গ্রহণসাধ্য তথ্য অনুসন্ধান করতে সহায়তা করে। এটি দেশের প্রধান বাংলা ও ইংরেজি পত্রিকাগুলো ছাড়াও বাংলা ব্লগ, বাংলা উইকিপিডিয়া এবং সরকারি তথ্য স্বয়ংক্রিয়ভাবে বিশ্লেষণ ও সংরক্ষণ করে। তাই সার্চ ইঞ্জিন ভিজিট করলেই দেখা যাবে এখানে রয়েছে চারটি আলাদা ক্যাটাগরির তথ্যানুসন্ধানের সুবিধা। এগুলো হচ্ছে– সংবাদ অনুসন্ধান, ব্লগ অনুসন্ধান, বাংলা উইকিপিডিয়া অনুসন্ধান ও জাতীয় ই-তথ্যকোষ।

সাধারণভাবে কেউ সার্চ করলে সংবাদ অনুসন্ধানের ফলাফল দেখানো হয়। সংবাদ ব্যতীত অন্যান্য অনুসন্ধান করতে হলে সার্চ বক্সের নিচে বাম দিকে চাহিদা অনুসারে ব্লগ/উইকিপিডিয়া অথবা ই-তথ্যকোষ অনুসন্ধানে ক্লিক করতে হবে।

সংবাদ অনুসন্ধান : পিপীলিকার সংবাদ অনুসন্ধান বিভাগটি আবার সাধারণ সার্চ, স্থানভিত্তিক সার্চ ও শ্রেণীভিত্তিক সার্চ এই তিনটি বিশেষ ভাগে বিভক্ত। স্থানভিত্তিক ও ক্যাটাগরিভিত্তিক সার্চ এখন শুধু বাংলার জন্য উন্মুক্ত।

সাধারণ সার্চ : সাধারণ সার্চে যেকোনো শব্দ/শব্দাবলী দিয়ে সার্চ করলে সেই শব্দের/শব্দাবলীর ভিত্তিতে সার্চের ফলাফল দেখানো হয়। যদি কেউ ইংরেজিতে সার্চ করে তবে কোনো ক্যাটাগরি অনুযায়ী ফলাফল দেয়া হয় না। সংবাদ সার্চের ফলাফলের পাশাপাশি একই শব্দ/শব্দাবলীর প্রাপ্যতা অনুযায়ী যেকোনো সার্চের প্রথম ১০টি ফলাফল দেখানো হয়। প্রয়োজনে তার নিচের আরও ফলাফল স্থানটিতে ক্লিক করে আরও অনুসন্ধান করা যায়।

পিপীলিকার বাংলা সার্চের জন্য আমাদের নিজস্ব একটি বাংলা অভিধান ব্যবহার করা হয়েছে। যদি ব্যবহারকারী কোনো শব্দের ভুল বানান দেন তাহলে পিপীলিকা স্বয়ংক্রিয়ভাবে সঠিক বানান খুঁজে নিয়ে সেই নতুন শব্দ দিয়ে অনুসন্ধান চালায়, ফলাফল দেয় এবং সাথে সাথে ব্যবহারকারীকে জানিয়ে দেয় তার কোন শব্দের বানান ভুল ছিল এবং সঠিক কোন শব্দ দিয়ে অনুসন্ধান চালানো হয়েছে। ব্যবহারকারী চাইলে পরে সেই ভুল শব্দ দিয়েই আবার অনুসন্ধান চালাতে পারেন। ইংরেজি সার্চের ক্ষেত্রে অভিধানটি ব্যবহার করা হয়নি। এই শ্রেণীতে ফলাফলে মোট অনুসন্ধানের সময়, মোট ফলাফলের মাঝে প্রথম কতগুলো ফলাফল দেখান হলো, প্রতিটি ফলাফলের শিরোনাম (হেডলাইন বা টাইটেল), প্রতিটি ফলাফল থেকে হাইলাইটেড বা চুম্বক অংশ, সংবাদটির মূল সোর্সের ওয়েব লিঙ্ক ও ক্যাশ করা সংবাদটি দেখান হয়।

এর মাধ্যমে ব্যবহারকারী মূল ওয়েব লিঙ্ক বা হেডলাইনে ক্লিক করলে সহজেই মূল সংবাদ সোর্সে যেতে পারেন। আবার ক্যাশে ক্লিক করলে সে এখানেই একসাথে সংবাদটির হেডলাইন, সোর্স, মূল সংবাদ, তারিখ ও ওয়েব লিঙ্ক দেখতে পারবেন। এজন্য তার মূল সোর্সে যাওয়ার প্রয়োজন নেই।

স্থানভিত্তিক সার্চ : কোনো ব্যবহারকারী যদি সার্চ বক্সে কোনো জেলার নাম ইংরেজিতে লেখার চেষ্টা করেন তাহলে পিপীলিকা তাকে স্বয়ংক্রিয়ভাবে বাংলাতে ওই স্থানের নাম সাজেশন করে থাকে। ব্যবহারকারী যদি শুধু স্থানটির নাম দিয়ে অনুসন্ধান করেন তাহলে পিপীলিকা তার ফলাফল প্রকাশের সাধারণ পদ্ধতিটি পরিবর্তন করে একটি বিশেষ পদ্ধতিতে ফলাফল প্রকাশ করে। সে সেই জেলার সাম্প্রতিক সময়ের তথ্যগুলো বিশ্লেষণ করে কয়েকটি ক্যাটাগরিতে প্রকাশ করে (যেমন– অপরাধ, ব্যবসায়, বিজ্ঞান ও প্রযুক্তি, খেলাধুলা, কৃষিতথ্য ইত্যাদি) প্রতিটি ক্যাটাগরিতে প্রথম পাঁচটিসহ মোট ফলাফল সংখ্যা দেখানো হয়। কোনো ব্যবহারকারী আরও ফলাফল স্থানে ক্লিক করে সহজেই ওই জেলার ওই ক্যাটাগরির বাকি ফলাফল দেখতে পারবেন। স্থানভিত্তিক সার্চের সময় অভিধান প্রয়োগ করা হয়নি।

ক্যাটাগরিভিত্তিক সার্চ : পিপীলিকায় বর্তমানে ফলাফলের মোট ছয়টি ক্যাটাগরি রয়েছে। এগুলো হচ্ছে– দেশের খবর, আন্তর্জাতিক, ব্যবসায় বাণিজ্য, তথ্যপ্রযুক্তি, বিনোদন ও খেলাধুলা।

জাতীয় ই-তথ্যকোষ ছাড়া বাকি ক্যাটাগরিগুলো বিভিন্ন ধরনের সংবাদ নির্দেশ করে। কোনো ব্যবহারকারী যেকোনো একটি ক্যাটাগরিতে ক্লিক করলে তিনি তার সার্চ বক্সে দেয়া শব্দ/শব্দাবলী শুধু ওই ক্যাটাগরির সংবাদগুলো অনুসন্ধান করতে পারবেন। সাধারণ সার্চের সময় সব ক্যাটাগরির সংবাদের মাঝে অনুসন্ধান চালানো হতো। এই সার্চের ক্ষেত্রেও সাধারণ সার্চের মতো ফলাফল দেখান হচ্ছে। জাতীয় ই-তথ্যকোষ সম্পূর্ণ ভিন্ন একটি ক্যাটাগরি হওয়ায় এই ক্যাটাগরিতে মূলত জাতীয় ই-তথ্যকোষ/জ্ঞানকোষের তথ্যগুলো সংরক্ষণের চেষ্টা করা হয়েছে।

পিপীলিকার কুশীলব

পিপীলিকার প্রকল্প পরিচালক হিসেবে বিশ্ববিদ্যালয়ের কমপিউটার প্রকৌশল বিভাগের অধ্যাপক ড. মুহাম্মদ জাফর ইকবাল এবং মুখ্য গবেষক ও টিম লিডার হিসেবে কাজ করেছেন মো: রুহুল আমীন সজীব। সহযোগিতায় ছিল বেসরকারি মোবাইল সেবাদাতা প্রতিষ্ঠান ▸

গ্রামীণফোন আইটি বিভাগের পাঁচ সদস্যের একটি দল। জিপিআইটির আর্থিক সহায়তায় পিপীলিকাকে প্রাণ দিয়েছেন শাবিপ্রবির ১১ জন ডেভেলপার। এরা হলেন– মো: মহিউদ্দিন মিশু, মাহবুবুর রব তালহা, তৌহিদুল ইসলাম, সাজ্জাদুল হক, বাকের মো: আনাস, আসিফ মো: সামির, মধুসূদন চক্রবর্তী অপু, আমিষ পাল, ফরহাদ আহমেদ, মাকসুদ হোসাইন ও তালহা ইবনে ইমাম।

পিপীলিকা নিয়ে এর টিম লিডার মো: রুহুল আমীন সজিব বলেন, পিপীলিকা আমাদের স্বপ্নপূরণের প্রথম ধাপ। প্রতিনিয়ত এর বৈশিষ্ট্য ও ফাংশন আপডেটের কাজ চলছে। যদিও সর্বপ্রথম আমি নিজে সার্চ ইঞ্জিন তৈরির পরিকল্পনা মাথায় এনেছি। শুরুতে অনেক ডেভেলপার এই প্রজেক্টে কাজ করলেও অবশেষে শাহজালাল বিজ্ঞান ও প্রযুক্তি বিশ্ববিদ্যালয়ের কমপিউটার বিজ্ঞান ও প্রকৌশল বিভাগের ১১ জন মিলে পুরো প্রজেক্টটি সম্পন্ন করি।

সজিব আরও জানান, ইতোপূর্বে জনপ্রিয় সার্চ ইঞ্জিনগুলোর কোনোটিতেই বাংলাভাষার ওপর তেমন গুরুত্বারোপ করা হয়নি। তাই পিপীলিকায় বাংলা তথ্য বিশ্লেষণ ও অনুসন্ধানের ওপর গুরুত্ব দেয়ার চেষ্টা করা হয়েছে। পিপীলিকার বাংলা সার্চের জন্য আমাদের নিজস্ব একটি বাংলা অভিধান ব্যবহার করা হয়েছে। যদি ব্যবহারকারী কোনো শব্দের ভুল বানান দেন, তাহলে পিপীলিকা স্বয়ংক্রিয়ভাবে সঠিক বানান খুঁজে নিয়ে সেই নতুন শব্দ দিয়ে অনুসন্ধান চালায়, ফলাফল দেয় এবং সাথে সাথে ব্যবহারকারীকে জানিয়ে দেয় তার কোন শব্দের বানান ভুল ছিল, সঠিক কোন শব্দ দিয়ে অনুসন্ধান চালানো হয়েছে।

রুহুল আমীন সজীব জানান, পিপিলিকাডটকম সবার জন্য উন্মুক্ত হলেও বিভিন্ন সরকারি-বেসরকারি প্রতিষ্ঠানের অভ্যন্তরীণ কাজের জন্য বিশেষভাবে সার্চ ইঞ্জিন তৈরি করার মাধ্যমে আমরা আয় করার চিন্তাভাবনা করেছি। আপাতত শুধু তথ্য খোঁজার কাজে সার্চ ইঞ্জিনটি ব্যবহার করা যাবে। ভবিষ্যতে ছবি, ভিডিওসহ বিভিন্ন কনটেন্ট খোঁজার সুবিধা যুক্ত করা হবে। তুলনামূলক চিত্র টেনে এনে গুগলে বাংলাভাষা ব্যবহার করা গেলেও বাংলা বানান শুদ্ধিকরণের কাজ কিন্তু তারা করে না যেটা পিপীলিকাডটকম করে থাকে। ফেজ ডিটেকশন বা পর্যবেক্ষণ, স্পেল চেক বা বানান শুদ্ধিকরণ, স্টেমার ও কীওয়ার্ড শনাক্তকরণ ফিচারগুলো আমাদের সার্চ ইঞ্জিনে আছে, যেগুলো গুগল সার্চ ইঞ্জিন থেকে পিপীলিকা সার্চ ইঞ্জিনকে আলাদা করে থাকে।

বাংলাভাষায় দেশের প্রথম পূর্ণাঙ্গ সার্চ ইঞ্জিন নিয়ে কথা হয় পিপীলিকার প্রকল্প পরিচালক মুহাম্মদ জাফর ইকবালের সাথে। তিনি জানান, আপাতত আমরা বাংলায় সার্চ ইঞ্জিন চালু করেছি মাত্র। আস্তে আস্তে এতে বিভিন্ন কনটেন্ট যোগ করব। আমাদের দীর্ঘ এক বছর লেগেছে পিপীলিকাডটকম দাঁড় করাতে। শাহজালাল বিজ্ঞান ও প্রযুক্তি বিশ্ববিদ্যালয়ের কমপিউটার বিজ্ঞান ও প্রকৌশল বিভাগের ১১ জন এবং জিপিআইটির পাঁচজন মিলে মূলত সার্চ ইঞ্জিনটি তৈরি করেছে। শাহজালাল বিজ্ঞান ও প্রযুক্তি বিশ্ববিদ্যালয়ের কমপিউটার বিজ্ঞান ও প্রকৌশল বিভাগের লেকচারার রুহুল আমীন সজীব পুরো প্রজেক্টের প্রধান হিসেবে কাজ করেছে। পিপীলিকাডটকম তৈরি করার সময় আমাদের মূল বিষয় ছিল সবকিছু হবে বাংলাভাষায়।

তিনি বলেন, আমরা এখন গর্ব করে বলতে পারব আমাদের ভাষার জন্য আমাদের নিজস্ব সার্চ ইঞ্জিন আছে। এটি সত্যিকার অর্থেই একটি সার্চ ইঞ্জিন, যার মাধ্যমে সারা বিশ্বের ২০ কোটি বাঙালি তাদের নিজের ভাষার তথ্য খুঁজতে পারবে।

এ সময় পিপীলিকায় তথ্য সংরক্ষণের ক্ষেত্রে অশ্লীল শব্দ যাচাই-বাছাই করার পাশাপাশি তা রাষ্ট্র, ধর্ম ইত্যাদিকে বিকৃত বা হেয় প্রতিপন্ন করে কিনা সে বিষয়েই নজর দেয়া হয় বলে জানান জাফর ইকবাল।

কথা প্রসঙ্গে তিনি বলেন, আমাদের পাবলিক বিশ্ববিদ্যালয়ের জন্য কোনো বড় তহবিল নেই। আমরা চাইলেও অনেক কিছু করতে পারি না। জিপিআইটি আমাদের এই ছোট প্রজেক্টটি অনেক বড় করে তৈরি করতে সাহায্য করেছে। আমি তাদের কাছে কৃতজ্ঞ।

পিপীলিকার চলার পথকে কুসুমাস্তীর্ণ করতে এগিয়ে আসা বিশ্বমানের দেশীয় প্রযুক্তি প্রতিষ্ঠান জিপিআইটির প্রধান নির্বাহী রায়হান শামসি জানান, এক বছরের অক্লান্ত চেষ্টার পর আমরা পিপীলিকাডটকম তৈরি করেছি। জাফর ইকবাল স্যারের সহযোগিতায় আমরা এই কঠিন কাজ করতে সক্ষম হয়েছি। আশা করছি অল্প কিছুদিন পর ডেস্কটপ ছাড়াও মোবাইল ব্রাউজার দিয়ে পিপীলিকাডটকম ব্যবহার করা যাবে। ভবিষ্যতে যেনো অ্যান্ড্রয়িড, উইন্ডোজ, আইওএস, ব্ল্যাকবেরি অপারেটিং সিস্টেমের সেলফোন থেকেও স্বাচ্ছন্দ্যে পিপীলিকা ব্যবহার করা যায় এজন্য মোবাইল অ্যাপ্লিকেশন তৈরির কথাও মাথায় রেখেছি।

পিপীলিকার এগিয়ে চলার সপ্তাহ পার হওয়ার আগেই পিপীলিকার জন্য একটি আন-অফিসিয়ালি ওপেনসার্চ প্লাগ-ইন তৈরি হয়েছে। এটি তৈরি করেছেন অনিরুদ্ধ অধিকারী। pipilika.adhikary.net ঠিকানায় প্রবেশ করে এই প্লাগ-ইনটি ডাউনলোড করা যবে। প্লাগ-ইনটি ইনস্টল করে চালু করলেই গুগল, ইয়াহু বা বিং-এর মতো ব্রাউজারের সার্চবার থেকেই ব্যবহারকারীরা ওয়েবসাইটে না গিয়েও পিপীলিকায় সরাসরি সার্চ করতে পারবেন কজ

ফিডব্যাক : netdut@gmail.com