

# Distribúované vyhľadávanie (DrPS)

Bc. Milan Laslop

Bc. Tomáš Hurban

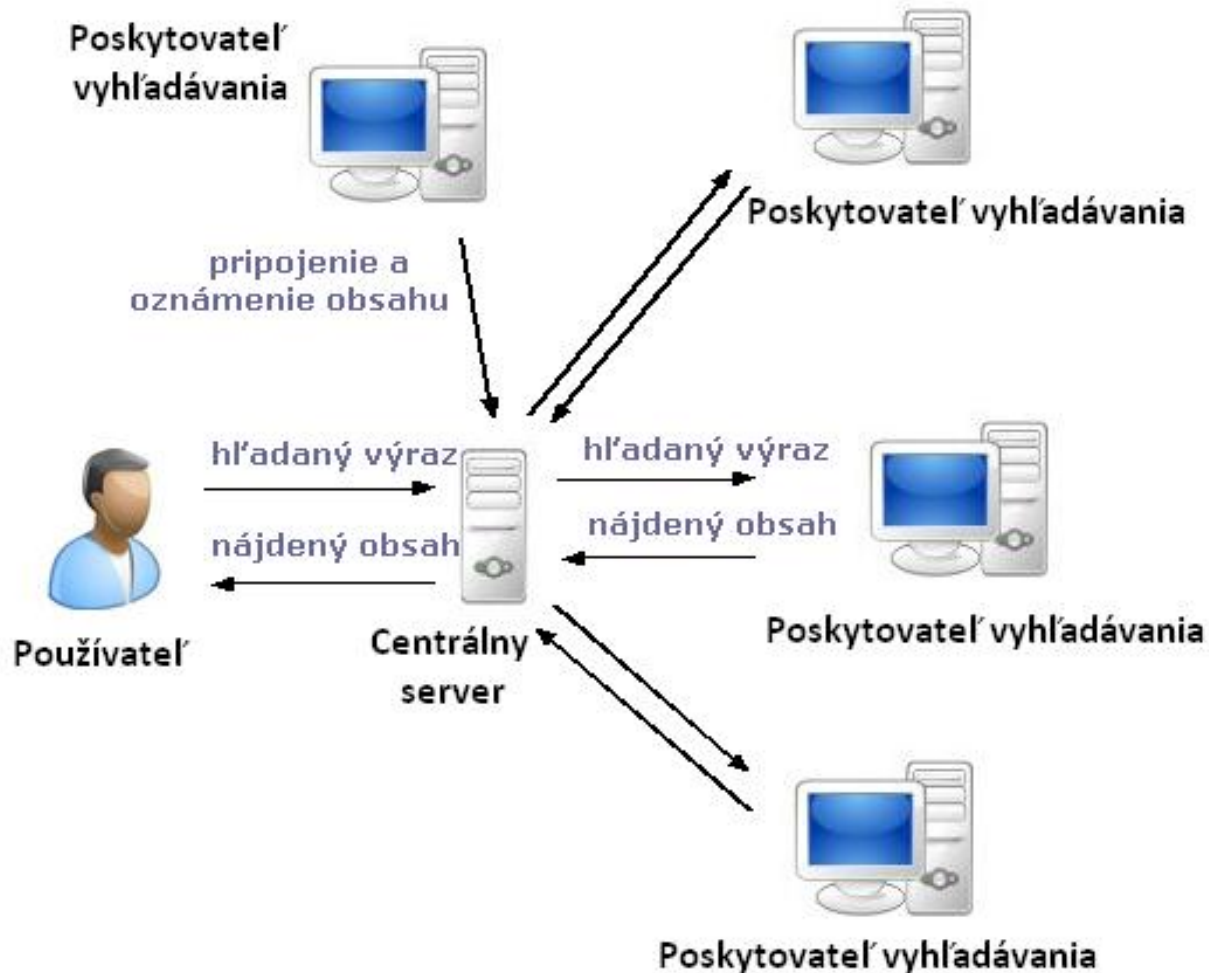
# Zadanie

- \* paralelné prehľadávanie zadaného obsahu
- \* update a synchronizácia nových častí obsahu

# Architektúra systému

- \* **Klient (používateľ)**
- \* **Centrálny server**
  - \* distribúcia obsahu
  - \* vyberanie vhodných poskytovateľov vyhľadávania
  - \* vyskladanie odpovede pre klienta z čiastočných odpovedí od poskytovateľov vyhľadávania
- \* **Poskytovateľ vyhľadávania**
  - \* vyhľadávanie v pridelenej časti obsahu

# Architektúra systému (2)



# Proces vyhľadávania

1. Klient zadá hľadaný výraz.
2. Server pošle požiadavky jednotlivým pripojeným poskytovateľom vyhľadávania na základe obsahu v ktorom môžu vyhľadávať.
3. Poskytovatelia obsahu vrátia nájdené texty.
4. Server tieto čiastočné výsledky poskladá do finálnej odpovede klientovi.

# Riešené distribuované problémy

- \* **Distribúovanie obsahu** – server oznamuje poskytovateľom vyhľadávania po ich pripojení, ktoré časti obsahu majú mať a tým zabezpečuje rovnomerné distribuovanie a zálohovanie obsahu.

# Riešené distribuované problémy (2)

- \* **Update starých častí obsahu** – centrálny server vie, ktorý poskytovateľ vyhľadávania má uložené ktoré časti obsahu, a preto pri zmene jednej časti obsahu vie poslať informáciu len tým poskytovateľom, ktorí túto časť obsahu majú, aby si mohli novú časť obsahu stiahnuť.

# Riešené distribuované problémy (3)

- \* **Rýchle vyhľadávanie vo veľkom obsahu** – centrálny server rozdelí celý obsah medzi poskytovateľov vyhľadávania a tým umožní paralelné prehľadávanie obsahu.



# Riešené distribuované problémy (4)

- \* **Riešenie výpadku poskytovateľa** – ak sa poskytovateľ odpojí od servera počas vyhľadávania zavolá sa funkcia *disconnect()* a tým sa vyhodí zo zoznamu dostupných poskytovateľov vyhľadávania.

# Nedoriešené problémy

- \* **Optimalizované vyhľadávanie** – server si pamätá, ktorému poskytovateľovi koľko trvalo vyhľadávanie v danej časti a snaží sa vyberať najrýchlejších poskytovateľov. Zaznamenávanie časov implementované je, no nie je implementované rozhodovanie o výbere najlepších poskytovateľov podľa týchto časov.

# Neriešené problémy

- \* **Riešenie výpadku centrálného servera** – ak sa stane, že spadne centrálny server, tak spadne celý systém, pretože používateľ sa nemá inak ako cez centrálny server dostať k výsledkom vyhľadávania.

# Neriešené problémy (2)

- \* **Synchronizácia viacerých centrálnych serverov bežiacich súčasne** – systém je navrhnutý na beh len s jedným centrálnym serverom.