

## 1. lecke

### Google keresési technikák

#### Bevezetés

A világháló legismertebb keresőszolgáltatása a Google. A Google keresés optimalizálását a legegyszerűbben a bővített keresési lehetőségekre (Advanced Search) kattintva is elvégezhetjük. Itt megadhatjuk többek között az alábbi feltételeket:

a keresett lapok nyelve,  
 a keresett lapok formátuma,  
 a keresett lapok frissítésének dátuma,  
 a keresett kifejezés előfordulása a lapon,  
 valamint kiszűrhetünk különböző tartományokat.

Lehetőség van azonban arra is, hogy néhány egyszerű keresési parancs szöveges beírásával, továbbá ezek (együtt)működési elvének megértésével a keresésünk eredményességét elősegítsük.

#### Alapvető tudnivalók

A legfontosabb az eredményes kereséshez, ha az intelligenciát a keresési folyamat során magunktól és nem a Google-től várjuk el. Akkor lesz hatékony a keresőkifejezésünk, ha nem kérdésként fogalmazzuk meg, hanem a várt találatból indulunk ki. Ha szövegre keresünk rá, a várt eredmény első szavait kell valamiképp kiötleni. Ha például könnyen szeretnénk milliárdosok lenni a helyes kifejezés nem az, hogy „hogyan nyerhetek a lottón?” hanem „A lottón úgy lehet nyerni,”.

Mindenképp célszerű minél rövidebben, két-három szóval megfogalmazni keresőkifejezésünket, vagy hosszabb keresés esetén felosztani több tagmondatra, több két-három szavas kifejezésre. A Google keresőben legfeljebb 32 szóra lehet keresni egyszerre. Ugyanakkor – a Boole-operátorok kivételével – nem befolyásolja a keresés eredményét, ha a kulcsszavakat kis- vagy nagybetűkkel írjuk, mivel a Google minden betűt kisbetűsként értelmez. Az alapértelmezett keresés az ékezeteket és megkülönböztető jeleket sem veszi figyelembe. Például a „Töhötöm" és a "Toehoetoem" elvileg ugyanazokat a keresési eredményeket adja.

A **plusz jelet** (+) a tiltott szavak aktiválásához használhatjuk. Az online könyvtári katalógusokhoz hasonlóan ugyanis elvileg tiltott szavaknak (*stop words*) számítanak a Google rendszerében is a keresés szempontjából érdemi jelentéssel nem bíró szavak. Ilyenek például angol nyelvi beállítások esetén: I, a, the, and, of. Ezeket úgy vonhatjuk be a keresésbe, ha közvetlenül eléjük + jelet teszünk, ez esetben a beírt kifejezés minden szavát figyelembe veszi. Például: +the Thames.

Más keresőkkel ellentétben a Google-ban semmilyen **helyettesítő karakterrel** (\*; ?; #) nem lehet pusztán egy karaktert helyettesítő keresést végrehajtani. A Google-ban végzett keresés esetén a **csillagot** (\*) abban az esetben használhatjuk, ha egy vagy több egész, teljes szót kívánunk helyettesíteni vele. Például: „piros \* zöld” vagy „Ég a \* \* szik \*”.

Ha megismétlünk egy keresőszót, azzal szűkíthetjük, és egyben módosíthatjuk a találati listát. Ismételni csak kétszer van értelme, a többszörös ismétlés érdemben elvileg már nem befolyásolja a találati listát, bár a Google minden keresésnél kulcsszó-sűrűség vizsgálatot is végez, azaz az alapján is rangsorol, hogy a keresett szó vagy kifejezés hányszor fordul elő a vizsgált weblapon.

A keresésünk felépítésénél a **mínusz jelet** (-) is használhatjuk. Ebben az esetben a Google mindazon oldalakat kihagyja a találati listából, amelyekben a mínusz jel után begépett kulcsszó szerepel.

Több keresőszóból kötött kifejezést idézőjelek segítségével alkothatunk. Például: „mohácsi csata”.

Az idézőjel használata azért is javasolt, mert nélküle sok irreleváns találat kapnánk. Az idézőjel különösen akkor lehet hasznos, ha például konkrét személynévre keresünk rá.

Keresésnél az is érdemes figyelembe venni, hogy a Google több szempont szerint súlyozza a találatait, tehát nem a legújabbak és adott esetben a legjobbak kerülnek az első helyekre, hanem a legnépszerűbbek, a legismertebbek. A Google az egyes találatok sorba rendezésénél az adott találat **PageRank** értéke (oldal-érték) alapján súlyoz. A művelet lényege, hogy a Google egy az A weboldaltól a B weboldalra mutató linket a B oldal jelentőségére vonatkozó értéknek könyvel el. Minél több link irányul tehát egy oldalra más weboldalokról, annál magasabbra kerül ez az oldal a Google-rangsorban. Magyarul a PageRank annak az összetett értéke, hogy egy adott oldalra hány különböző link és honnan mutat.

Míg a PageRank egy oldal népszerűsége alapján, addig a Google úgynevezett **geotargeting** tulajdonsága a földrajzi-nyelvi azonosítótól függően rendezi sorba a találatokat. Ez azt jelenti, hogy mindig a helyi információkat helyezi előre, ami gyakran a találatok nem kívánt szürkítését eredményezheti.

~~Ha teljes lehetséges találati listában szeretnénk böngészni, érdemesebb a **google.com/ner** (ner = no country region/redirect) URL-t a címsorba gépelve indítani a keresésünket. Célszerű azért is az angol nyelvű felületről indulni, mivel a Google-nak nem teljes értékűek a honosított változatai, így a magyar nyelvű se biztosít azonnali hozzáférést a keresőóriás által biztosított összes opcióhoz. 2016 óta nem működik .~~

A Google lehetőséget biztosít úgynevezett kategóriakeresésre, amit a <http://directory.google.com> cím alatt kezdeményezhetünk. Ez tulajdonképpen a könyvtárak tárgyszó-katalógusaihoz hasonlóan weboldalak ezreit kategorizálja önkéntes segítők munkájára és a PageRank technológiára épülve. A keresőmező segítségével az egyes alkategóriákon/tárgyszavakon belül lehetséges a szokásos Google keresés is. Adott esetben így hatékonyabb lehet a keresés, mint az egyszerű webkeresésnél.

## ***A Boole-operátorok használata Google keresésnél***

A Boole-operátorok a Google kereséseknél is eredményesen használhatóak. A legalapvetőbb operátor több keresőszó együttes megadása esetén a szóköz, amely alapértelmezés szerint a keresőszavak, illetve kifejezések között az **AND** (és) relációt fejezi ki. Például: Mohács csata 1526 azt jelent: Mohács AND csata AND 1526. Ebben az esetben értelemszerűen csak azok a weblapok szerepelnek találatként, ahol mind a három megadott szó előfordul.

A keresés találati listáját megváltoztathatjuk, ha úgy írjuk be a keresőablakba, hogy

Mohács OR csata OR 1526, ekkor a kereső külön keres rá mind a három szóra. Az **OR** helyett állhat ez a jel is: | (AltGr+w).

Természetesen lehet kombinálni is a keresőparancsokat. Például: (Mohács OR csata) AND 1526 leütése után a kereső a Mohács vagy csata kulcsszavakat tartalmazó honlapokat és az 1526-ot tartalmazó találatokat listázza.

Fontos: a Boole-operátorok csak akkor működnek, ha nagybetűvel írjuk őket.

## Műveletek parancsszavakkal

Az egyszerű keresés mellett a Google alkalmas arra is, hogy különböző parancsszavak, úgynevezett **Google kódok** (Google Code) vagy másképpen **prefixek** (előtagok) használatával első ránézésre talán bonyolultabb, ám egyben sokszorosan hatékonyabb keresést hajtsunk végre. Az alábbiakban a legfontosabb parancsszavak ismertetése következik.

Fontos: a kettőspont része a parancsszavaknak! Azt a keresőkifejezést, amelyre szeretnénk, hogy a Google kód hatással legyen, szóköz nélkül írjuk a parancsszó után.

| Google kód                          | Feladata  | Példa  |
|-------------------------------------|---|--|
| intitle:                            | Ezzel a weboldal címében (<title>) kereshetünk. Használható a Google speciális keresőszolgáltatásain is (Group, Images, News)   | intitle:szociológia<br>intitle:"index of"    |
| inurl:                              | A weboldalak URL-jében, honlapcímében keres, szótöredékre nem, csak egész elemre. A Web és az Images keresésnél is használható.<br>Az utolsó példának a segítségével megtudhatjuk, hogy pillanatnyilag hány weboldalt lát a Google, becslések szerint ez a szám a létező oldalak 25-50%-a között van. | inurl:news<br>inurl:belföld<br>inurl:http:// |
| intext:                             | A weboldalak szövegéből listáz, a linkeket, a címeket és URL-eket kihagyja.   | intext:"tanulási<br>segédletek"              |
| allintitle: allintext:<br>allinurl: | E kódváltozatok annyiban különböznek az eredeti parancsszótól, hogy hatásukra az összes a kettőspont után felsorolt keresőszónak szerepelni kell az egyes találatokban.   | allinurl:hir belföld                         |
| link:                               | Ez a megadott honlapra mutató linkeket listázza.  | link:oszk.hu                                 |
| cache:                              | Ezzel a paranccsal kizárólag a Google által az indexelés során eltárolt honlapok közt böngészhetünk, így hozzáférhetünk az esetlegesen már törölt oldalakhoz is, úgy, hogy közben megőrizzük anonimitásunk.   | Cache:"Otto von<br>Habsburg"                 |
| filetype:<br>ext:                   | Mindkét kóddal tetszőleges fájl típusra szűkíthetjük a keresésünk. A leghasznosabb a Web és a Groups kereséseknél.  | filetype:pdf<br>ext:doc                      |
| site:                               | Ezzel a paranccsal lehet a legegyszerűbben kizárólag egy adott oldalra, vagy egy adott ország területére lokalizált oldalakra rákeresni. Az első példa a kari honlap oldalain keres rá a  | site:btk.ppke.hu<br>történelem               |

|           |   |  |
|-----------|---|--|
|           | történelem kulcsszóra. A második pedig kizárólag de (német) tartományvégződésű oldalak körén belül keres.   | site:de  |
| inanchor: | Meglinkelt szavakra, kifejezésekre kereshetünk rá vele.   | inanchor:"Weöres Sándor"   |
| x..y      | Számsoron belül az adott két érték közé írt két ponttal kereshetünk. Ugyancsak a két pontot használhatjuk a minimum és a maximum érték megadásánál. | „Anglia története”<br>1200..1800<br>pendrive ..5000<br>oldalszám 100.. |
| ~         | Ha egy kulcsszó elé a hullámvonalat tesszük, akkor szinonimákat, hasonló szavakat is keres.   |  |
| related:  | A megadotthoz hasonló oldalakat kereshetünk, hatása ugyanaz, mint amikor a <i>Similar Pages</i> lehetőséggel élünk a találati listában.             |  |
| define:   | Egy adott fogalom meghatározását keresi. Angol nyelven működik egyelőre.  |  |
| info:     | a teljes URL megadása után ad információkat az adott weblapról  | info:bt.k.ppke.hu  |

Google search results for the query: `max weber ext:pdf site:de -intext:center inurl:uni`. The search returned approximately 29,700 results in 0.18 seconds. The left sidebar shows search filters: Everything, Images, Videos, News, Shopping, and More. The main results list includes:

- Religion & Economic Growth and Development** (PDF) from [mpra.ub.uni-muenchen.de/8181/1/MPRA\\_paper\\_8181.pdf](http://mpra.ub.uni-muenchen.de/8181/1/MPRA_paper_8181.pdf). File Format: PDF/Adobe Acrobat - Quick View. by S Nath - 2007 - Cited by 2 - Related articles. **Max Weber**, an eminent sociologist of early 20 th century, in his seminal work, The Protestant Ethic and the Spirit of Capitalism, introduces the linkage ...
- Political Legitimacy** (PDF) from [edoc.ub.uni-muenchen.de/5813/1/Musonda\\_Anthony.pdf](http://edoc.ub.uni-muenchen.de/5813/1/Musonda_Anthony.pdf). File Format: PDF/Adobe Acrobat. by AM Musonda - 2006 - Related articles. **Max Weber's** Ideal Type Theory of the State. 16. 2 . The Origins of the State ...
- Max Weber - Wissenschaft als Beruf** (PDF) from [www.wsp-kultur.uni-bremen.de/.../Max%20Weber%20...](http://www.wsp-kultur.uni-bremen.de/.../Max%20Weber%20...). File Format: PDF/Adobe Acrobat - Quick View. **Max Weber**. WISSENSCHAFT ALS BERUF. 1919. Ich soll nach Ihrem Wunsch über » Wissenschaft als Beruf« sprechen. Nun ist es eine gewisse Pedanterie von uns ...
- Max Weber als Nationalökonom und Wirtschaftshistoriker** (PDF) from [www.wiwi.uni-frankfurt.de/profs/schefold/docs/Sem07SS%20Weber.pdf](http://www.wiwi.uni-frankfurt.de/profs/schefold/docs/Sem07SS%20Weber.pdf). File Format: PDF/Adobe Acrobat - Quick View. 1. Lehrstuhl für Volkswirtschaftslehre, insb. Wirtschaftstheorie. Prof. Dr. Dres h.c. Bertram Schefold. **Max Weber** als Nationalökonom und ...
- Welfare State Institutions and Welfare State Outcomes** (PDF) from [www.mzes.uni-mannheim.de/.../EQUALSOC\\_Traleq\\_Workshop\\_MZES\\_...](http://www.mzes.uni-mannheim.de/.../EQUALSOC_Traleq_Workshop_MZES_...). File Format: PDF/Adobe Acrobat - Quick View. Dr. Barbara M. P. Lechner: **Revisions of the Max Weber Thesis in the 19th and 20th century...**

Kombinált keresés a Google-n.

A keresést a fenti kódok tetszőleges és értelemszerű kombinációival tovább lehet szűkíteni. Például: `filetype:pdf intext:tematika` kombinált paranccsal olyan pdf találatokat kapunk, amelyek szövegében szerepel a tematika szó.

Kombinálásra alkalmatan Google kódok: `link:`, `info:`, `related:`, `allintext:`, `allinurl:`, `allintitle:`.

| Kombinálásra alkalmatlan esetek  | Példa  |
|--|--|
| ha a két kitétel kizárja egymást,  | <code>site:hirado.info -inurl:hirado</code>                                      |
| ha a kitételek összeférhetetlenek, ilyen helyzetben az OR operátorral lehet helyessé tenni a kifejezést, | <code>Erasmus site:de site:hu</code><br><code>Erasmus (site:de   site:hu)</code> |

## Google Hacking

A keresők által nem látott területet mélywebnek (rejtett web, láthatatlan web, deep web) nevezzük, ennek mérete egyes becslések szerint több százszorosa a kereshető (surface) webnek.

A mélyweb olyan oldalakat tartalmazhat, ahol

- 1) az adott oldal tiltja a keresőrobotnak a hozzáférést,
- 2) az oldal dinamikusan adatbázisból a felhasználó kérésére jön létre (például egy könyvtári katalógus,
- 3) regisztrációhoz kötött az oldal megtekintése,
- 4) speciális, a kereső által értelmezhetetlen formátumban készült oldal (Flash),
- 5) elszigetelt site, nem mutat rá külső link.

A Google kódokkal azonban olyan, különböző kombinációkat is létre lehet hozni, amelyek a világháló nehezen hozzáférhető részeihez, esetleg éppen a mélyweb alapesetben zárt tartalmihoz nyithatnak utat a tapasztalt felhasználók számára. Az ilyen műveleteket összefoglaló néven **Google Hacking**-nek nevezzük. A Google Hacking fogalmát és a ráépülő adatbázist (**Google Hacking Database**) egy amerikai hacker és számítógépes biztonsági szakember, Johnny Long dolgozta ki. Az általa kezelt nyilvános adatbázisban számos ilyen kombinációt találhatunk magyarázatokkal.

## A Google URL változtatása

Google keresésünket nem csak parancsszavakkal, hanem úgy is hatékonyabbá tehetjük, ha módosítjuk a Google találati listájának a címsorban található URL-t. Ez az URL alapesetben sem nevezhető egyszerűnek, értelemszerűen, ha a keresőkifejezésünk bonyolultabb, a Google kereső URL-je még összetettebb lesz. Néhány elem hozzátoldásával, illetve módosításával, azonban könnyedén befolyásolhatjuk a találati listánkat.

A fontosabb URL elemek magyarázata

| URL elem           | Magyarázat   |
|--------------------|--|
| <code>&amp;</code> | az elemek közti elválasztó jel   |
| <code>num=</code>  | az oldalanként megjelenő találatok száma (ez csak akkor működik, ha a beállításoknál az automatikus kiegészítést kikapcsoljuk) |

|             |  |
|-------------|--|
| lr=         | a találati oldalak nyelvét állíthatjuk be vele   |
| hl=         | a keresőfelület által használt nyelvet mutatja   |
| Restrict=   | kiválasztott országra, régióra szűkíti a keresést  |
| q=<br>as_q= | a keresőkifejezés áll utána  |
| as_qdr=     | az oldal frissítésének dátuma alapján szűkít. Például az as_qdr=m4 a négy hónapnál frissebb oldalak közül tallóz, míg a d30 az utóbbi 30 napból. |

A restrict, a hl és az lr nyelvi változókkal befolyásolhatjuk keresésünket. Ha ugyanis ezeket megváltoztatjuk az URL-ben, akkor az a további keresésekre, illetve a keresőoldalunk felületére is hatással lesz. A hatékony keresés érdekében fontos, hogy a többféle nyelvi beállítást vagy a Google tulajdonságainál (Preferences) vagy közvetlenül az URL keresztül biztosan kézben tartsa a felhasználó. A nyelvi kódok könnyen megjegyezhetőek, általában az adott ország domain végződésével azonosak. Néhány példa:

| lr kód     | hatása  |                        |
|------------|---------|------------------------|
| lr=lang_fr | francia | nyelvű oldalakon keres |
| lr=lang_de | német   |                        |
| lr=lang_hu | magyar  |                        |

| hl kód | hatása  |                                 |
|--------|---------|---------------------------------|
| hl=fr  | francia | nyelvű lesz a böngésző felülete |
| hl=hu  | magyar  |                                 |
| hl=la  | latin   |                                 |

| restrict kód       | hatása     |   |
|--------------------|------------|---|
| restrict=countryVA | vatikáni   | domain végződés alatt bejegyzett, illetve az adott országban található szerverszolgáltató által tárolt oldalak közül ad találatot |
| restrict=countryAT | ausztriai  |   |
| restrict=countrySK | szlovákiai |   |

## **A Google környezetre gyakorolt hatása**

A Google hasznos segédeszköz, de csak ha megfontoltan használjuk. A megfontoltság nemcsak a kapott találatok ellenőrzésénél és értékelésénél javasolt, hanem környezetvédelmi szempontok miatt is. Alex Wissner-Gross, amerikai fizikus kutatásokat folytatott annak érdekében, hogy kiderítse, mekkora környezeti hatása van a Google-keresésnek. A kutató tanulmánya rámutat, hogy egy tipikus Google-keresés közel 7 gramm szén-dioxidot állít elő. A második keresés esetén ez a szám már a duplájára növekszik, ez a mennyiség pedig már megegyezik egy elektromos teafőző által generált szén-dioxiddal. Wissner-Gross kutatása rámutatott arra is, hogy 0,02 gramm szén-dioxidot hozunk létre minden egyes online töltött percben. Ez egyáltalán

nem hangzik soknak, viszont szakértők napi 200 millióra teszik az internetes keresések számát. Mindebből az következik, hogy globálisan az IT-szektor annyi üvegházhatású gázt termelt, mint az összes légitársaság együttesen.

**Tipp**

Érdemes a Google-n kívül más keresővel, vagy metakeresővel is próbát tenni. Felmérések szerint a különböző keresőszolgáltatások találatai összességében csupán 10-15%-os egyezést mutatnak. Példák metakeresőre:

<http://www.dogpile.com/>

<https://www.ixquick.com/>

<http://www.deeperweb.com>

**Kérdések, feladatok**

1. Melyik Google kód kombináció helytelen? Miért? Javítsa is ki a hibásakat!
  - a. `intext:"erdelyi örmények" ext:pdf -site:com`
  - b. `digitale inurl:uni allintitle:heidelberg site:de`
  - c. `Ibn Khaldún site:hu site:com ext:pdf`
  - d. `Therion -inurl:book -intitle:book ext:jpg`

**Szakirodalom**

Calishain, Tara–Dornfest, Rael: Google Hacks.Pub.: O'Reilly. ISBN : 0-596-00857-0 több kiadás

Henk van Ess: Der Google-Code: Die Geheimnisse der besten Suchergebnisse. Addison-Wesley, München 2011.

Long, Johnny: Google Hacking for Penetration Testers, Volume 2. Syngress Publishing, Inc. Elsevier, Inc. 2008.

Google Hacking Database <http://www.exploit-db.com/google-dorks/>