

Ide nekem az internetet – wget

A szélessávú netkapcsolatok elterjedésének hála, ma már ezerféle anyaghoz – szoftverekhez, dokumentációkhoz, képekhez, videókhoz – hozzájuthatunk, csak le kell tölteni őket. Böngészőnkben általában elég egy kattintás és a kiválasztott anyag máris töltődik le gépünkre. Ez azonban csak igen korlátozott lehetőségeket nyújt számunkra.

■ Ha egy valóban kezes segítő társat szeretnénk letöltéseinkhez, használjuk a *wget*-et. Nagyon hatékony segédeszközünk lehet hálózatról való letöltésre *ftp* és *http* (vagy *https*) protokollokon keresztül. Régi segítő társunk ő, bár lehet, hogy a szekrény mélyén lapult, így vegyük elő, töröljük le róla a port és ismerkedjünk meg vele!

A *wget* minden *Linux* és *BSD* terjesztésnek része, de *MacOS X*-en és *Windows*-on is használhatjuk, sőt létezik *VMS*-re írt változata is. A *wget* karakteres felületen vagy terminálban használható, tehát ismernünk kell a kapcsolókat, de mint látni fogjuk, ez igen hasznosnak fog bizonyulni mindennapos munkánk vagy egy-egy komolyabb letöltés során.

A wget képességeinek áttekintése

Az internet alapvetően nem megbízható közeg, így ha nagyobb anyagot, vagy sok kicsit kell letöltenünk, gyakran kell bosszankodnunk a folyamat megszakadása miatt. Ha pedig egy honlapot mentenénk le minden kellékével együtt, néha a böngészőnk nem elég intelligensen oldja ezt meg ahhoz, hogy a tükrözött helyet lokálisan is jól tudjuk majd használni. A *wget* minderre egyszerű megoldást nyújt. Hogy lássuk, mennyire hatékony eszközről van szó, tekintsük át röviden a program fontosabb tulajdonságait, képességeit.

- Szinte bármilyen **nix* környezetben megtalálható, vagy lefordítható, sőt szabadon használhatjuk *Windows* vagy *MacOS X* alatt is.
 - Képes a háttérben dolgozni, így csak akkor kell odafigyelünk rá oda, amikor szeretnénk, akár ki is jelentkezhethetünk.
 - Lassú, vagy nem megfelelő kapcsolat esetén egy bizonyos határon belül addig próbálkozik, míg a letöltés jó nem lesz.
 - Tudja folytatni a megszakadt letöltéseket (*ftp* kapcsolat esetén persze csak akkor, ha ezt a kiszolgáló is támogatja).
 - Proxy kiszolgálóval is használhatjuk a letöltés gyorsítására, a hálózati terhelés csökkentésére.
 - Teljes könyvtárrendszereket letölthetünk vele a távoli gépről.
 - Letöltési célként megadhatunk fájl típusokat, vagy meghatározhatjuk a kihagyni kívánt csoportot is.
 - Konvertálthatjuk vele a linkeket a helyi gépre, azaz automatikusan kialakíthatjuk az „offline böngészésre” alkalmas anyagot.
 - Megszabhatjuk, hogy a linkek követése során mennyire „kalandozhat el” más helyekre, honlapokra.
 - A letöltés során nyomon követhetjük a folyamat menetét, de mindezt át is irányíthatjuk egy fájlba, ahol később tanulmányozhatjuk.
 - Minden beállítást elmenthetünk egy konfigurációs fájlba, így nem kell minden egyes alkalommal megadnunk őket. Ezeket egyedileg felül is bírálhatjuk, ha éppen erre van szükségünk.
 - Végül, de nem utolsó sorban teljesen szabad, nyílt forráskódú szoftverről van szó, ami jelentős érv a *wget* mellett a rengeteg érdekes, de a legkevésbé sem olcsó letöltésvezérlők korában.
- Azt hiszem, ennyi információ épp elég annak eldöntéséhez, hogy szeretnénk-e közelebről megismerni a *wget*-tel...
Mivel valószínűleg megtaláljuk kedvenc terjesztésünkben, ezért letölteni sem szükséges. Ha mégis szeretnénk a legfrissebb változathoz hozzájutni (ez jelenleg, 2005. július elején az 1.9.1-es), akkor a <http://www.gnu.org/order/ftp.html> oldalon található ftp szerverek listájáról válasszunk egyet.

Alapvető tudnivalók

Ennyi bevezető után vágjunk a közepébe! A használat nagyon egyszerű:

```
wget [opciók] [URL]
```

Amennyiben az opciókat kezdésként elhagyjuk, a kérdés az marad, mi is pontosan az „URL”? Ez egy hálózati erőforrás azonosítója, ami lehet dokumentum, kép, hang vagy bármilyen más fájl. Az **URL** rövidítés az „**Uniform Resource Locator**”-t takarja, amit talán „Egységes erőforrás-azonosítónak” magyarosíthatnánk, az internetes erőforrások szabványosított címe, tartalmazza a protokollt (például HTTP), a használt portszámot, a gazdagép nevét és az erőforrás elérési útját a gépen belül, esetleg még az a név és jelszó-párost is, ami az eléréshez szükséges. Nézzünk egy példát!

A `http://hu.wikipedia.org` például egy nagyon egyszerű **URL**, tulajdonképpen csak egy protokollt, a HTTP-t tartalmazza és az internetcímet. (A magyar **Wikipédiát** a `hu.wikipedia.org` címen találjuk, ha az információkra éhes olvasó esetleg még nem ismerné, akkor sürgősen keresse fel.)

Az `ftp://pisti:pst32z@pistinet.hu/fajlok/doksi.zip` már bonyolultabb. Itt már megadtuk a nevet és a jelszót is, ami az ftp eléréshez szükséges (név: `pisti`, jelszó: `pst32z`) és a teljes elérési utat a „`pistinet.hu`” szerveren belül. Megadom a `wget` számára értelmezhető **URL** mintát általános módon is. Ez a következő:

```
protokol1://név:jelszó@gazdagép
↳:port/könyvtár/fájl
```

A protokoll lehet `http`, `https` vagy `ftp`. Ennek megadása kötelező. A név és jelszó kettősponttal van elválasztva, amit akkor kell csak megadnunk, ha azonosítanunk kell magunkat. Névtelenül is használható (anonymous) `ftp` szerverek esetén ezt a `wget` elvégzi helyettünk. A név és jelszó megadását kukac jellel zárjuk le, ezután megadjuk a gazdagép domén nevét vagy **IP** címét, ezután kettősponttal elválasztva a portszámot, ha az nem a szabványos (a szabványos portszámok: 80-as

`http` esetén, 443 `https`-nél és 21 `ftp` szolgáltatásnál) majd az elérendő erőforrás útvonala következik.

Most már tudjuk, hogy az **URL**-ek meglehetősen bonyolultak is lehetnek, viszont a jó hír az, hogy szerencsére ritkán kell a szabványostól eltérő porton lévő szerverre felcsatlakoznunk, publikus szervereknél pedig a nevet és jelszót sem kell megadnunk. Persze amikor egy hosszú **URL**-t kell megadnunk, könnyen eltéveszthetjük. Ezért érdemes a címet a böngészőből kimásolni, hiszen az esetek többségében valószínűleg az interneten bukkanunk rá egy-egy izgalmas helyre, ahonnan azonnal le szeretnénk valami tölteni. A `wget` igazán akkor kezd érdekessé válni, ha megismerkedünk az opciókkal. Ezek segítségével szabályozhatjuk, hogyan töltsön le, mit vegyen figyelembe és mit nem és miként jelenítse meg a letöltés folyamatát. Ne feledkezzünk meg arról sem, hogy a kis- és nagybetűk bizony számítanak az opciók megadásakor, így például a „`V`” és „`v`” egészen mást jelentenek.

Alapvetően három lehetőségünk van:

- `v`: Ez a verziószámot írja ki, ami hasznos lehet megállapítani, hogy vajon a legújabb verziót használjuk-e vagy pedig letöltőnk már frissítésre szorul
- `h`: A `wget` helpjét (magyarosan a súgóját) kapjuk meg
- `b`: Megadásakor a `wget` a háttérben fog futni, mindenféle visszajelzés nélkül. Ez nagyon hasznos lehet, ha a terminálban tovább szeretnénk dolgozni. Ami normálisan a képernyőre kerülne, az most egy „`wget.log`” nevű fájlba íródik bele. Így a letöltés után végigkövethetjük, mi történt, felderíthetjük az esetleges hibákat is.

Ha be akarom mutatni a további lehetőségeket is, akkor az opciókat csoportosítanom kell, mert nagyon sok van, hiszen ahogy említettem, a `wget` nagyon rugalmas szoftver, szinte minden, a letöltéssel kapcsolatos óhajunkat képes teljesíteni.

Én négy csoportra osztottam fel a funkciók listáját, bár nyilván lehetne máshogyan is:

- A ki- és bemenet szabályozása (megjelenítési opciók, naplózás, **URL** listafájl megadása)
- A letöltés szabályozása (proxy szerver megadásától a letöltés folytatásának lehetőségén át a letöltés sebességéig sok minden befér ide)
- **HTTP** opciók (a `http` protokollon keresztüli letöltés szabályozása)
- **FTP** opciók (az `ftp` protokollon keresztüli letöltés szabályozása)

A `wget` opciók általában az url megadása előtt következnek, de semmi sem kötelez erre bennünket – ha úgy jobban tetszik, egyeseket átcsoportosíthatunk utána is. A következőkben tehát az opciókkal foglalkozom, de nem fogom mindet felsorolni, inkább csak bemutatni a `wget` lehetőségeit.

A ki- és bemenet szabályozása

```
--progress = típus
```

A „típus” helyére írt szöveggel megadhatjuk, milyen legyen a letöltés folyamatának a kijelzése. Ha nem használjuk ezt a beállítást, akkor az alapértelmezett mintát fogjuk látni, ami jelenleg a bar. Ez a következőképpen fest:

```
32% [=====>                ]
↳14,968,576    118.56k/s
↳ETA 04:22
```

Az első szám a letöltött mennyiséget mutatja százalékban, ezután két kapcsos zárójel között (ez a bar, azaz „téglalap” egysoros ASCII formája) egy folytonosan növekvő nyílát láthatunk. A zárójel után a letöltött mennyiség bájtokban, majd az aktuális sebesség (kilobájt per másodpercben) és az ETA látható. Az ETA az „*Estimated Time of Arrival*”, azaz a „becsült várakozási idő” rövidítése. Használhatjuk még dot-ot is, ennek további altípusai is vannak. Írhatjuk önmagában is, ami az alapértelme-

zett „dot” stílust fogja jelenteni, amikor is egy pont 1 kilobájtot fog jelenteni, minden sorba pedig 50 pontot ír ki. Ez nyilván a régi, lassú hálózatok maradványa, amikor még csak kis dolgokat tölthettünk le belátható időn belül, így ezzel kényelmesen nyomon követhettük a letöltés folyamatát. Egy szélessávú hálózaton ez elviselhetetlenül szétaprózza a kijelzést. Az „elavult” jelzést ráakaszthatjuk a dot:binary típusra is, ami érdekes lehet, az a dot:mega, ahol névével ellentétben nem egy megabájt egy pont, hanem 64K. Minden sorba 48 pontot ír, így egy sor 3 megabájtnak fog megfelelni:

```
wget -progress=dot:mega
↳ ftp://ftp.uhu.linux.hu/ISO/
↳ uhu-linux-1.2-cd1.iso
```

Ez valahogy így fog kinézni (1. kód) A dot stílus kiválóan alkalmas arra, hogy nyomon kövessük, az egyes egységekben (itt ez 3 megabájt) mekkora volt az átlagos letöltési sebesség. A bar-ral ellentétben sokkal finomabb egységekben jelzi a letöltés folyamatát is. Ez főleg egyenetlen letöltési sebességnél jöhet jól nekünk.

-o logfile

A *wget* alapértelmezésben a képernyőre ír mindent. Ezt átírányíthatjuk egy fájlba ezzel az opcióval. A -b opciónál említettem hogy itt automatikusan a *wget.log* fájlba kerül a program kimenete. Az -a opció csak annyiban különbözik az -o-tól, hogy amennyiben létezik a megadott fájl, nem törli azt, hanem folytatja.

A *wget*-et teljesen el is „hallgattathatjuk” a -q vagy --quiet opcióval. Ezzel ellentétben a -v vagy --verbose opció „szószátyárrá” teszi a *wget*-et, ami egyébként az alapbeállítás is. A kettő között van az -nv vagy --non-verbose opció, ami egy általában hallgatag, de a hibáknál „beszélő” szoftvert eredményez.

Előfordul, hogy több fájl kell letöltenünk, amik nem egy helyen vannak (mert akkor a rekurzív letöltést alkalmazhatjuk) hanem az interneten szétszórva, több gazdagépen. Ilyenkor alkalmazhatjuk a -i file vagy -input-file = file opciót, amelynél egy megadott fájlból fogja venni a letöltendő elemeket.

A letöltés szabályozása

Ha megszakad a letöltés és a távoli szerver ezt engedi, a *wget* tudja folytatni azt, erre szolgál a -c opció. Amennyiben a hálózati kapcsolat szakad meg és nem a *wget* működése, ez az alapértelmezése, tehát automatikusan folytatni fogja a már megkezdett letöltést. Előfordulhat azonban, hogy kézzel állítjuk le a *wget*-et vagy valamilyen más hiba történik, ekkor az opció megadása nélkül újból elkezd letölteni a fájlt, meghagyva az eredetileg félbehagyottat is.

Nagyon hasznos funkció lehet a letöltés sebességét szabályozó --limit-rate opció. Ezzel beállíthatjuk, mekkora az a legnagyobb sebesség, amivel a *wget* letölthet. Így ha közben szeretnénk internetezni vagy a belső hálózaton lévő más felhasználóktól nem szeretnénk a teljes sáv szélességet elvenni, ezzel szabályozhatjuk a nekünk tet-sző sebességet.

Bájt, kilobájt és megabájt per másodperc mértékegységeket használhatunk. Tehát a

```
wget --limit-rate=35k
↳ ftp://ftp.fsn.hu/pub/linux/
↳ distributions/knoppix/
↳ KNOPPIX_V3.9-2005-05-27-EN.iso
```

azt jelenti, hogy maximum 35 kilobájt/másodperc, azaz 280 kbit/s sávveléssel fog menni a *Knoppix* disztribúció letöltése. Így ha például egy 512 kbit/s adsl vonalunk van, akkor a többi felhasználónak vagy más letöltéshez még mindig marad 232 kbit/s.

Ha *FTP* kiszolgálóról töltünk le, használhatjuk a helyettesítő karaktereket is:

```
wget ftp://ftp.szerverem.hu/* .zip
```

De több könyvtár tartalmát is letölthetjük a -r (rekurzív letöltés) opcióval. Ez nemcsak *FTP*-n, hanem weboldalakon is alkalmazható. Ha ugyanis enélkül akarunk egy honlapot letölteni:

```
wget http://www.linuxvilag.hu
```

akkor csupán egy *index.html*-t fog letölteni, ugyanis ezt kapta vissza a *wget* a *Linuxvilág* web szerverétől. A honlap további részeinek letöltéséhez a

```
wget -r
↳ http://www.linuxvilag.hu
```

formát kell használnunk.

Ez már egy *www.linuxvilag.hu* nevű könyvtárat is létrehoz és ebbe menti le a honlap tartalmát. Web oldalon ez az egyes linkeket is megvizsgálja, így a letöltés ezekre is kiterjed. Alapértelmezésben 5 szint mélységig fog letölteni, de ezt az -l (vagy --level=) opcióval megváltoztathatjuk. Például a -l 2 megadásával két szint mélységig fog vizsgálni. Végtelen mélységet a -l 0 opcióval adhatunk meg, de ezzel nagyon vigyázzunk, mert lehet hogy bizonyos helyekről több gigabájt adatot fogunk letölteni.

Ha teljes tükrözést szeretnénk, ezt az -m opcióval érhetjük el (ekkor az -r és -l nem kellene).

A weboldalak letöltésének további lehetőségeiről a következő fejezetben lesz szó.

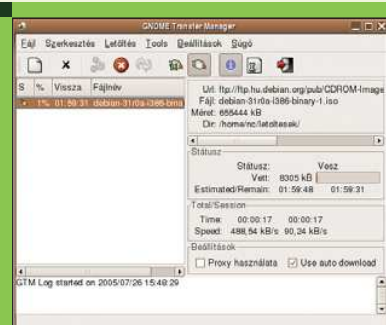
Végül, érdemes megemlíteni a következő három opciót:

```
-nc (vagy --no-clobber)
-N (vagy --timestamping)
-K (vagy --backup-converted)
```

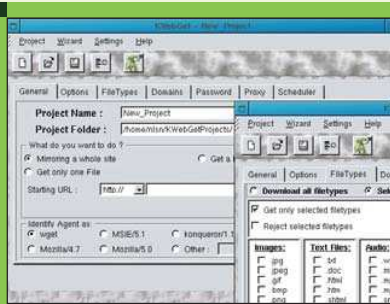
Ezek a rendszeres letöltésnél használhatók jól. Ha például egy webhelyet rendszeresen frissítünk a gépünkön, fontos a -N opciót használni, így csak a változásokat fogja letölteni. Ha ezt a -K -val együtt

1. kód

```
OK ..... 0% 112.62 KB/s
3072K ..... 1% 114.58 KB/s
6144K .....
```



■ 1. ábra A Gnome Transfer Manager



■ 2. ábra KwebGet



■ 3. ábra wGetGUI

használjuk, akkor a régi fájlokról egy biztonsági mentést (backup) is készít, azaz átnevezi őket, hogy ne kerüljenek felülírásra. A „no clobber” opció (-nc) viszont megakadályozza, hogy a létező fájlokat felülírassuk, így ezek le sem töltődnek. Ez akkor hasznos, ha mindig csak az új fájlokat szeretnénk egy helyről letölteni (és az új fájlok megnevezése is eltérő a régiekéktől).

HTTP opciók, weboldalak letöltése

A weboldalak letöltéséhez számtalan lehetőséget, beállítást kínál nekünk a *wget*. Mint említettem, ha nem adunk meg a letöltésre vonatkozó opciót, akkor csak az index oldalt fogjuk visszakapni. Rekurzív letöltés esetén (-r opció) viszont a megadott (vagy az alapértelmezett 5) szintig letölti a webtartalmat, követe a hivatkozásokat is (már amihez jogosultságunk van, természetesen *HTTP* protokoll esetén is megadhatunk bejelentkezési (login) nevet és jelszót akár az *URL*-ben, akár a *-http-user=user* és *--http-password=password* opciókat használva). A letöltött oldalak és webtartalom ezután böngészőben megnézhető... lenne, ha a hivatkozások a helyi gépre mutatnának. De alapértelmezésben mindent úgy töltünk le, ahogy a webhelyen található. Ekkor csak a relatív címzések fognak működni, az abszolútak nem. Persze lehet, hogy nem szeretnénk az eredeti tartalmat megváltoztatni. Viszont az is előfordulhat, hogy a weboldalon egyes linkek más helyekre is mutatnak, a *wget* így végül akár a szándékunkkal eltérő tartalmakat is letölthet vagy a letöltés a végtelenségig folytatódik.

Ha offline böngészésre szeretnénk használni a letöltött website tartalmat, mindenképpen konvertálni kell a hivatkozásokat a -k (vagy --convert-links) opcióval. Ekkor minden link a helyi gépre fog mutatni. A *wget* szerencsénkre más tartományokba (domain) nem „kalandozik el”, azaz ha a letöltendő weboldalakon kívülre mutató link van, azt nem dolgozza fel. Tehát a

```
wget -r http://www.debian.org
```

csak a *debian.org* tartományra mutató linkeket fogja követni. Ellenben ha a -H opciót is megadjuk, akkor minden linket követni fog, bárhová is hivatkozzon. A -D opcióval tudjuk korlátozni ezt a funkciót, megadva, mely tartományokra korlátozzuk a letöltést. Tehát a

```
wget -rH -Ddebian.org
↳ http://www.debian.org
```

letölti a *debian.org* altartományai lévő tartalmat is, így a *packages.debian.org*-ot vagy a *lists.debian.org*-ot is. Természetesen csak a legritkább esetben szeretnénk akkora tartalmakat letölteni, mint a *packages.debian.org*. Ezért a nem kívánt helyek listáját megadhatjuk a --exclude-domains opcióval:

```
wget -rH -Ddebian.org
↳ --exclude-domains
↳ packages.debian.org,lists.
↳ debian.org http://www.debian.org
```

Az is egy jó módszer a letöltés kordában tartására, ha csak a relatív linkeket engedjük követni a *wgetnek* a -L opcióval. A *www.debian.org* tipikusan

ilyen jól megszerkesztett oldal, ahol relatív linkek mutatnak a tartományon belüli tartalomra. Végül, a webes letöltések bemutatását fejezzük be egy nagyon hasznos lehetőséggel, a *tükrözéssel (mirroring)*. Ez természetesen a -m opcióval adható meg. Egy webhely (vagy akár *FTP* hely) tükrözése során a *wget* rekurzív módon mindent letölt az adott tartományból (természetesen *HTTP* esetén a linkek követésével), ami a legutóbbi tükrözés óta megváltozott.

FTP opciók

FTP szerverről letöltésnél a leggyakoribb opciónk valószínűleg a megfelelő név és jelszó megadása lesz:

```
--ftp-user=user
--ftp-password=password
```

Ezek hiányában a *wget* *anonymousként* próbál belépni. Használhatjuk a speciális behelyettesítő karaktereket is: a * vagy ? szimbólumokat.

```
wget ftp://szerverem.hu/*.jpg
```

Így természetesen csak a *jpeg* képeket fogja letölteni számunkra. Általában passzív módban töltünk le *FTP* szerverekről. De előfordulhat, hogy csak az aktív mód engedélyezett, ekkor kell használnunk a --no-passive-ftp opciót. Gyakran alkalmaznak szimbolikus linkeket az *FTP* szerverek könyvtárstruktúrájában. A *wget* alapértelmezés szerint ezeket nem követi, tehát gépünkre is csak a link kerül. Ha szeretnénk, hogy a szimbolikus linkek által

hivatkozott helyekről is letöltsön, alkalmaznunk kell a `--retr-symlink` opciót.

A `wgetrc`

A `wgetrc` egy nagyon hasznos beállítófájl, ahol a `wget` általános működését tudjuk szabályozni. Több helyen is megtalálhatjuk. A globális beállítások az `/etc/wgetrc` fájlban találhatóak (egykes disztribúciókban ez az `/usr/local/etc/wgetrc`). A felhasználótól függő beállításokhoz a saját könyvárunkban kell létrehozni egy `.wgetrc` fájlt. Ez felülbíráhatja az általános `wgetrc` beállításait.

Az egyes opciók megadása természetesen bármely `wgetrc`-t felülbírá. Ezek a beállítások addig érvényesek, míg a parancsokban felül nem bírájuk őket. A `wgetrc` szintaxisa nagyon egyszerű: változó = érték

Minden egyes változó-érték párt új sorba kell írni.

Rengeteg lehetőségünk van a `wget` szabályozására. Ha belenézünk a globális `wgetrc`-be, láthatjuk, hogy egy pár érték

már előre be van állítva számunkra, amik így a `wget` alapértelmezései. Ha használunk proxy szerveret, itt egyszer s mindenkorra megadhatjuk azt, elkerülve hogy minden letöltéskor be kelljen írni:

```
http_proxy = server1.ceg.hu
ftp_proxy = server2.ceg.hu
proxy_user = "janos"
proxy_password = "jelszavam"
```

A http://ftp.gnu.org/pub/gnu/Manuals/wget-1.8.1/html_node/wget_27.html#SEC27 oldalon egy teljes listát kaphatunk a `wgetrc` parancsokról.

Grafikus felületek

A `wget` használatának megkönnyítésére grafikus felületek is készültek hozzá. A *Gnome Transfer Manager* (1. ábra) egyszerű, grafikus letöltésmenedzser a *Gnome* ablakkezelőhöz, és természetesen a `wget`-en alapul. Ennél részletesebb beállítási lehetőségeket kínál *Frank von Daak KwebGet*-je (<http://www.kpage.de/en/index.html>) (2. ábra).

A `wget` windowsos változatát is letölthetjük a <http://gnuwin32.sourceforge.net/packages/wget.htm> oldalról. Ezen alapul a *wGetGUI* (<http://www.jensroesner.de/wgetgui/>) is (3. ábra), amellyel nagyon részletesen szabályozhatjuk a `wget` számtalan lehetőségét.

Egyetlen cikkben természetesen lehetetlen bemutatni a `wget` minden tulajdonságát és szolgáltatását, tehát a sikeres felfedezésekhez bátran használjuk a `man wget` parancsot vagy a www.gnu.org/software/wget/manual/ oldalon található részletes ismertetőt.



Molnár Norbert

34 éves, rendszergazdaként dolgozik, 5 éve foglalkozik Linuxszal. Főként a szabad szoftverek és a számítógépes biztonság érdekli. Budapesten él, hobbija – ha éppen nem egy újabb disztribúciót tanulmányoz és nem utazik – a csillagászat és a filozófia – lehetőleg jó vörösbor mellett. (molnar.norbert@gmail.com)

