

OBĚD ZDARMA SE SHANNONEM

1.1 KÓD ISBN

Vezměte do ruky jakoukoliv novější knihu. Na zadní desce nebo v tiráži najdete číslo ISBN, neboli International Standard Book Number (Mezinárodní standardní číslo knihy). ISBN jednoznačně označuje knižní titul mezi všemi knihami vydanými na celém světě. Desetimístné ISBN této knihy je*

80-7363-289-6.

(Pomlčky nejsou pro naše potřeby důležité a zachoval jsem je, aby se číslo snadněji četlo.) Toto číslo má překvapivou vlastnost. Podívejme se na tabulku 1.1 – v prvním sloupci jsou číslice ISBN, ve druhém čísla od 1 do 10 a třetí tvoří řádkové součiny obou předchozích sloupců. Například ve třetí

ISBN	Koeficient	Součin
8	1	8
0	2	0
7	3	21
3	4	12
6	5	30
3	6	18
2	7	14
8	8	64
9	9	81
6	10	60
Součet		308 = 11 × 28

Tabulka 1.1. Zvláštní vlastnost kódu ISBN

* Zde uvedený desetimístný kód se nedávno změnil na 13-místný, princip však zůstal stejný.

1. OBĚD ZDARMA SE SHANNONEM

tím řádku jsou první dvě číslice 7 a 3, a tak třetí je $7 \times 3 = 21$. Když je třetí sloupec hotov, sečteme všech 10 čísel v něm a součet, 308, napíšeme dolů.

Možná jste si všimli, že součet, 308, je dělitelný jedenácti. To není náhoda; stalo by se to, ať byste si vzali jakoukoliv knížku. Zkuste to sami s nějakou svou oblíbenou knihou. (Jestli ISBN končí na X, zacházejte s X jako s číslem 10.)

Možná vás nejdřív napadlo, že to je jeden z těch matematických kouzelnických kousků; ať napíše jakékoliv číslo, bude takto vytvořený součet vždy dělitelný jedenácti. Abychom viděli, že to tak není, zkusme změnit poslední číslici ISBN z 6 na 7. Dolní člen třetího sloupce se změní z 60 na 70 a součet se zvětší o 10, na 318, což jedenácti dělitelné není. Za dělitelnost součtu jedenácti tedy vděčí naše ISBN nějaké své zvláštnosti. Mají ji všechna čísla ISBN a otázkou je, proč. Proč se asi nakladatelé rozhodli používat jen určitá zvláštní čísla k označování svých knih? Představte si, že si od nakladatele objednáváte knihu. Vyplníte ISBN do objednávacího formuláře, ale uděláte malou chybičku; na páté místo omylem napíšete 7 místo 6. Co to udělá se součtem ve 3. sloupci? Pátý člen 3. sloupce se změní z 30 na 35. Tím se součet zvětší o 5 a přestane být dělitelný jedenácti. Když nakladatel obdrží vaši objednávku, hned vidí, že jste si objednali knihu, jejíž identifikátor nepatří k těm správným zvláštním číslům – žádná kniha s takovým číslem neexistuje. Proto ví, že jste udělali chybu, a místo aby vám poslal nesprávnou knihu, což by se při dnešním poštovním dost prodražilo, požádá vás o opravu objednávky.

ISBN je příkladem samodetekujícího kódu. Nejen že v něm jsou zakódovány údaje o knize (vydavatel, název atd.), ale navíc dokáže samočinně objevovat chyby; jestli náhodou zaměníte nějakou číslici, je adresátovi hned jasné, že došlo k omylu. Totéž se stane, když zaměníte dvě sousední číslice (což se lidem často stává). Než budete číst dál, zkuste si vyřešit následující úlohu.

Úloha 1.1. Dokážete objevit, která vlastnost kódu mu dává schopnost odhalit záměnu číslic?

Odpověď na tuto otázku spočívá ve druhém sloupci uvedené tabulky. ISBN dokáže záměnu objevit zásluhou jednotlivých násobitelů 1, 2, ..., 10 ve druhém sloupci. Třeba mějte za sebou dvě číslice a a b , řekněme na 4. a 5. místě. Pak 4. a 5. řádek tabulky vypadá takto:

ISBN	Koeficient	Součin
\vdots	\vdots	\vdots
a	4	$4a$
b	5	$5b$
\vdots	\vdots	\vdots

Kdybyste teď vyměnili a s b , budou příslušné řádky vypadat takto:

ISBN	Koeficient	Součin
\vdots	\vdots	\vdots
b	4	$4b$
a	5	$5a$
\vdots	\vdots	\vdots

Součet 3. řádku by pak vzrostl o a a zmenšil se o b , ježto příspěvek čísla a se změní ze $4a$ na $5a$ a příspěvek od b z $5b$ na $4b$. Tím se součet změní o $a - b$. Nový součet nemůže být dělitelný jedenácti, není-li $a - b$ nula (ve kterémžto případě jsou a a b tytéž číslice a jejich záměnou se nic nezmění).

Použití násobků k rozlišení příspěvku různých číslic v ISBN je hlavním důvodem, proč je kód založen na dělitelnosti jedenácti a nikoliv deseti, což by se mohlo zdát přirozenější.* Na rozdíl od čísla 11 totiž není 10 prvočíslo. Kdybychom třeba použili dělitelnosti číslem 10, zůstal by některým číslicím „Černý Petr“. Řekněme, že 5. řádek vypadá takto:

ISBN	Koeficient	Součin
\vdots	\vdots	\vdots
3	5	15
\vdots	\vdots	\vdots

Jestli v tomto ISBN místo číslice 3 napíšete omylem 7, změní se číslo ve 3. sloupci z patnácti na 35 – tedy o 20. Ježto je 20 násobkem deseti, podmínka dělitelnosti desítkou by záměnu neodhalila. Potíž je v tom, že 10

* Kódy založené na prvočíslech mají i další výhody, ty jsou však pro potřeby ISBN nadbytečně jemné.

1. OBĚD ZDARMA SE SHANNONEM

je 5×2 , a tak když změníte v ISBN tuto číslici o 2, 4, 6 nebo 8, změníte číslo 3. sloupce o násobek deseti. U prvočísel, jako je 11, se to nemůže stát.

Když jsem se poprvé setkal se samoopravnými kódy, byl jsem úplně nadšen. Pouhou opatrnou volbou „jazyka“, v němž sdělujete zprávy, dokážete samočinně snížit pravděpodobnost nedorozumění. Jak si umíte představit, tato matematická myšlenka má v naší moderní době elektronického přenosu informací početná uplatnění. Ale samoopravné kódy jsou svým způsobem jedním z nejstarších lidských vynálezů. Všechny jazyky jsou vlastně kódy, které vybavují základní informaci dodatečnými mluvnickými a skladebními prostředky, jež umožňují posluchači ověřovat význam toho, co se říká. Tento příklad pomáhá objasnit, že samoopravné kódy nemají nic společného s kryptografií – s kódy, jež skrývají informace před nepřáteli. Smyslem samoopravného kódu je přenášet sdělení přesně, ne je přenášet tajně.

1.2 BINÁRNÍ KANÁLY

Když vesmírné sondy, jako ta vyobrazená na obrázku 1.1, posílají naměřené údaje (snímky Marsu či Jupiteru) zpět na Zem, musejí je nějak zakódovat, obvykle do posloupnosti nul a jedniček. Rádiové vlny, jež je přenášejí, musejí projít překážkou, představovanou zemským ovzduším, takže údaje, které tady dole pochyťáme, často nejsou tytéž, co byly poslány. *Kanál*, jímž se vzkaz předává, není dokonale přesný.

Jedním způsobem, jak tuto potíž obejít, by bylo nahradit všechny jedničky jejich delší posloupností (řekněme deseti jedničkami),

1111111111,

a všechny nuly deseti nulami,

0000000000.

Zachytí-li vaše pozemská stanice posloupnost

1101110111,

víte, že to není původně poslaná posloupnost, a můžete si být skoro jisti, že správná posloupnost je složena ze samých jedniček. Nemusíte tedy pochybovat, že správná značka je 1, a ne 0. Tento krok, nahrazení všech číslic dvojkové soustavy (neboli tzv. bitů) delší řadou bitů, vytváří kód, ač hrubý. Jako v kódu ISBN, určitá poselství jsou „nedovolená“, takže když nějaké



Obr. 1.1. Umělcova představa průzkumné sondy NASA k Marsu, připravené k vypuštění roku 2005.

takové obdržíte, poznáte, že došlo k chybě. Avšak tento kód vám poslouží o něco lépe než ISBN – umožňuje vám chybu *opravit*, a ne jen *objevit*. To je dobré, protože zatímco nakladatel vás může požádat o přeobjednání knihy, sondu nemůžete požádat, aby znovu vyfotografovala Mars, když už je na cestě k Jupiteru. Tento „opakovací“ kód, jak bychom jej mohli nazvat, je jednoduchým příkladem *samoopravného* kódu.

Potíž s tímto opakovacím kódem je v tom, že je velmi marnotratný. K poslání každého bitu informace musí sonda poslat 10 bitů údajů. Účinnost je jen 10%. Když se snažíte posbírat informace od vesmírné sondy, je důležité mít vysokou rychlost přenosu, protože sonda může každou planetu navštívit jen nakrátko a má jen omezené množství energie. To připomíná zásadu, že „nic není zadarmo“, jež v matematice vystupuje velmi často. Chcete-li vysoký stupeň přesnosti, musíte obětovat rychlost přenosu.