

Využití celulárních automatů pro kompresi dat

Katedra teoretické informatiky a matematické logiky

Bakalářská práce 2016

autor práce: Marek Polák

vedoucí práce: Mgr. Otakar Trunda



MATEMATICKO-FYZIKÁLNÍ
FAKULTA
Univerzita Karlova

Celulární automaty

Celulární automat (CA) je diskretní matematický model simulující vývoj konfigurace buněk v čase. Bývá chápán jako n -rozměrné pole buněk, kde každá buňka může nabývat m různých stavů. Stavem automatu se rozumí konkrétní konfigurace stavů jednotlivých buněk. Stav automatu se mění v čase po diskretních krocích – generacích. Stav každé buňky v generaci t je určen pomocí přechodové funkce ze stavů jejích sousedních buněk (včetně jí samé) v generaci $t-1$. Simulace celulárního automatu je tedy plně určena jeho počátečním stavem – generací 0.

Cíle práce

Celulární automaty jsou v informatice stále nedostatečně probádanou oblastí, přestože již bylo prokázáno mnoho jejich užitečných vlastností. I přes jejich velmi jednoduchou definici jsou například turingovsky úplně či schopné generovat vysoce náhodné binární posloupnosti z triviálních vstupních dat. Nabízí se tedy otázka, zda je možno využít celulární automaty k bezztrátové kompresi dat, resp. k transformaci dat vhodné pro jejich následnou kompresi.

Cílem této práce je prokázat či vyvrátit hypotézu, že v závislosti na parametrech mohou některé typy celulárních automatů transformovat vstupní data takovým způsobem, že tato se stanou snáze komprimovatelnými některým bezztrátovým kompresním algoritmem. Tento algoritmus může být jak již zavedený, tak vyvinutý speciálně za tímto účelem.

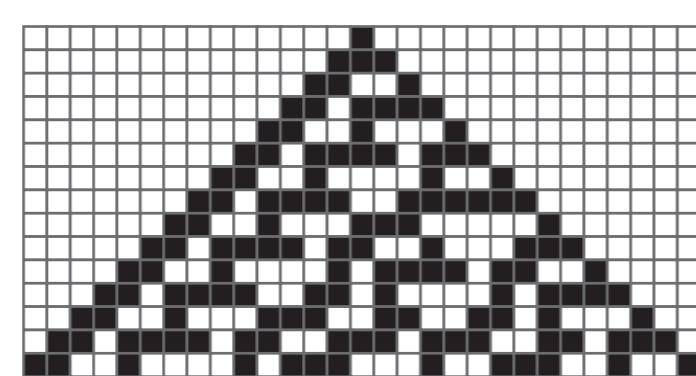
Práce se omezuje na zkoumání elementárních, tj. jednorozměrných dvoustavových celulárních automatů. Důležitým cílem tedy je klasifikovat elementární celulární automaty podle kritérií souvisejících s jejich využitelností k bezztrátové kompresi dat. Mezi tato kritéria patří např. počet unikátních dosažitelných stavů automatu z počátečního stavu či uspořádanost dosažitelných stavů. Za tímto účelem je implementováno též jednoduché uživatelské rozhraní pro vizualizaci a simulaci elementárních celulárních automatů.

V případě potvrzení hypotézy bude výsledkem práce bezztrátový kompresní algoritmus, využívající transformaci pomocí elementárního celulárního automatu a dosahující na reálných (textových a obrazových) datech kompresního poměru srovnatelného např. se slovníkovými algoritmy LZ77 či LZ78. Časová náročnost algoritmu není rozhodující, jedná se především o demonstrativní potvrzení počáteční hypotézy.

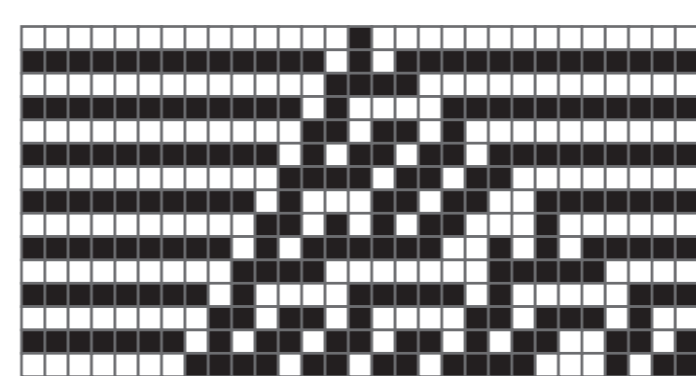
Klasifikace elementárních celulárních automatů

Elementární celulární automaty je možno klasifikovat na základě jejich přechodové funkce. Z tohoto hlediska existuje celkem 256 druhů elementárních CA, označovaných jako Wolframova pravidla. Každé z těchto pravidel může být dále aplikováno na nekonečném poli buněk nebo na poli buněk s pevnou velikostí, které může být chápáno cyklicky (krajní buňky spolu sousedí) či acyklicky. Tato práce se zabývá pouze automaty s konečnou velikostí.

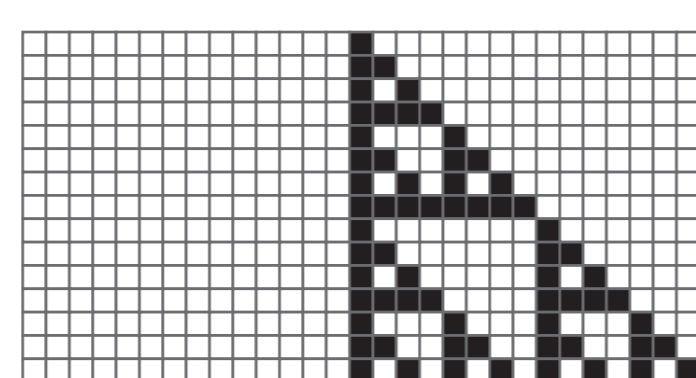
Diagramy vpravo znázorňují prvních 15 generací vybraných Wolframových pravidel z počátečního stavu s jedinou živou buňkou (první řádek).



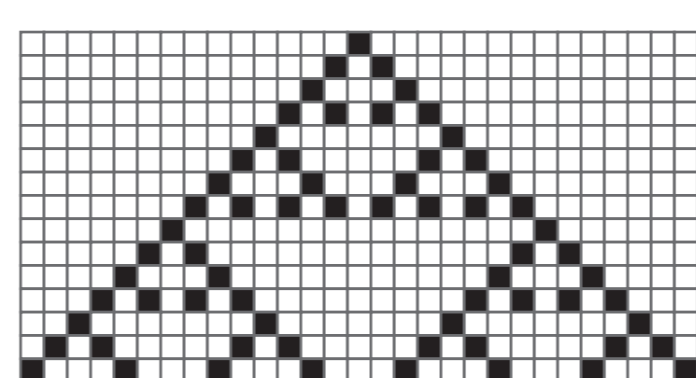
Wolframovo pravidlo 30



Wolframovo pravidlo 45



Wolframovo pravidlo 60



Wolframovo pravidlo 90

Kompresní a dekompresní algoritmus Cell

KOMPRESSE:

- (1) přečti blok délky n bitů ze vstupu
- (2) použij získanou bitovou sekvenci jako počáteční stav elementárního CA
- (3) simuluj zpětný běh CA do nalezení nejlépe komprimovatelného stavu
- (4) zapiš generaci, ve které byl stav získán, na výstup
- (5) komprimuj získaný stav vhodným kompresním algoritmem
- (6) zapiš komprimovaný stav na výstup

DEKOMPRESSE:

- (1) přečti ze vstupu číslo t
- (2) přečti ze vstupu počet bitů, který bude dekomprimován na n -bitovou sekvenci
- (3) dekomprimuj přečtenou bitovou sekvenci
- (4) použij získanou sekvenci jako počáteční stav elementárního CA
- (5) simuluj t generací daného CA
- (6) zapiš bity získaného stavu na výstup

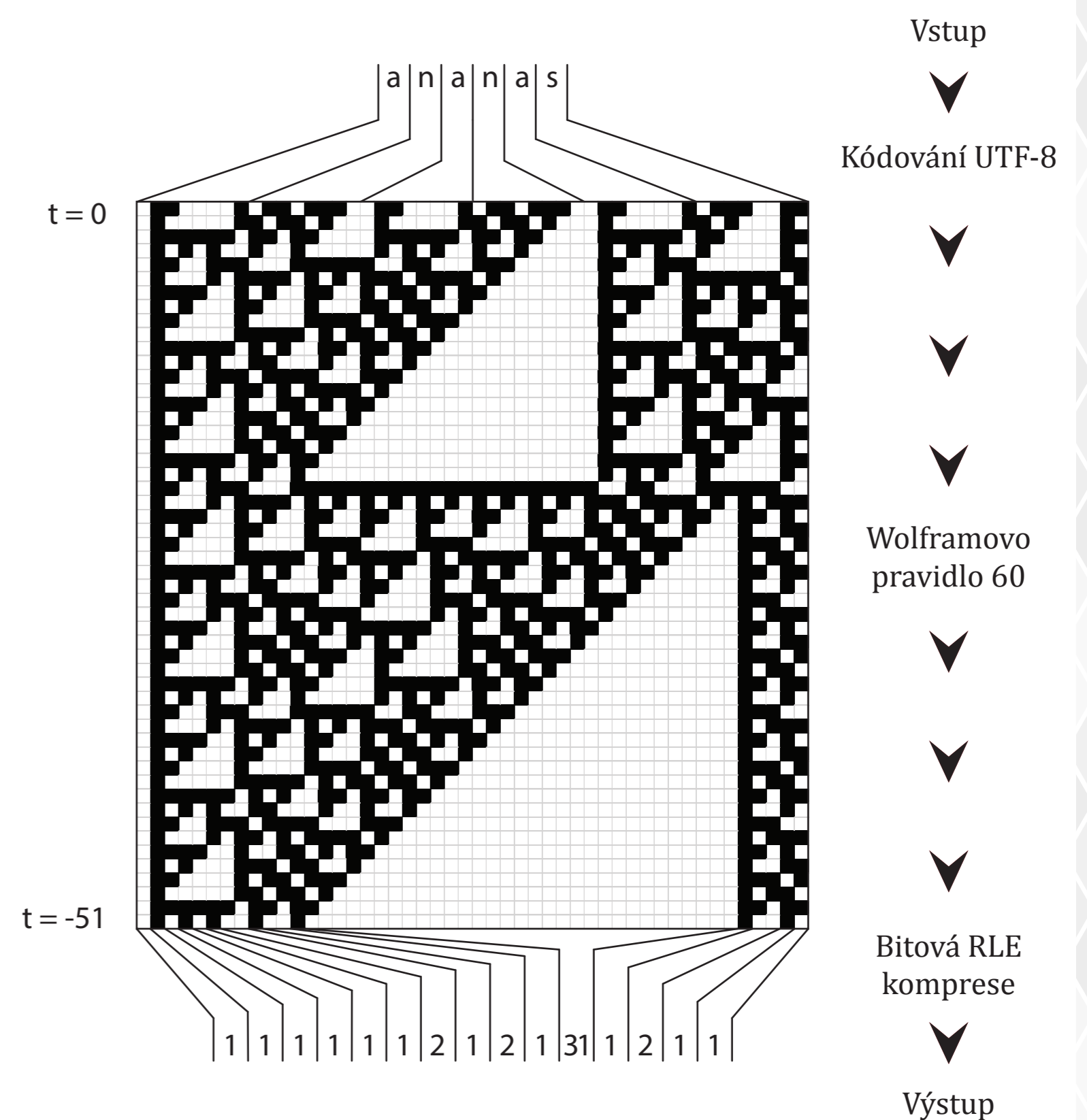
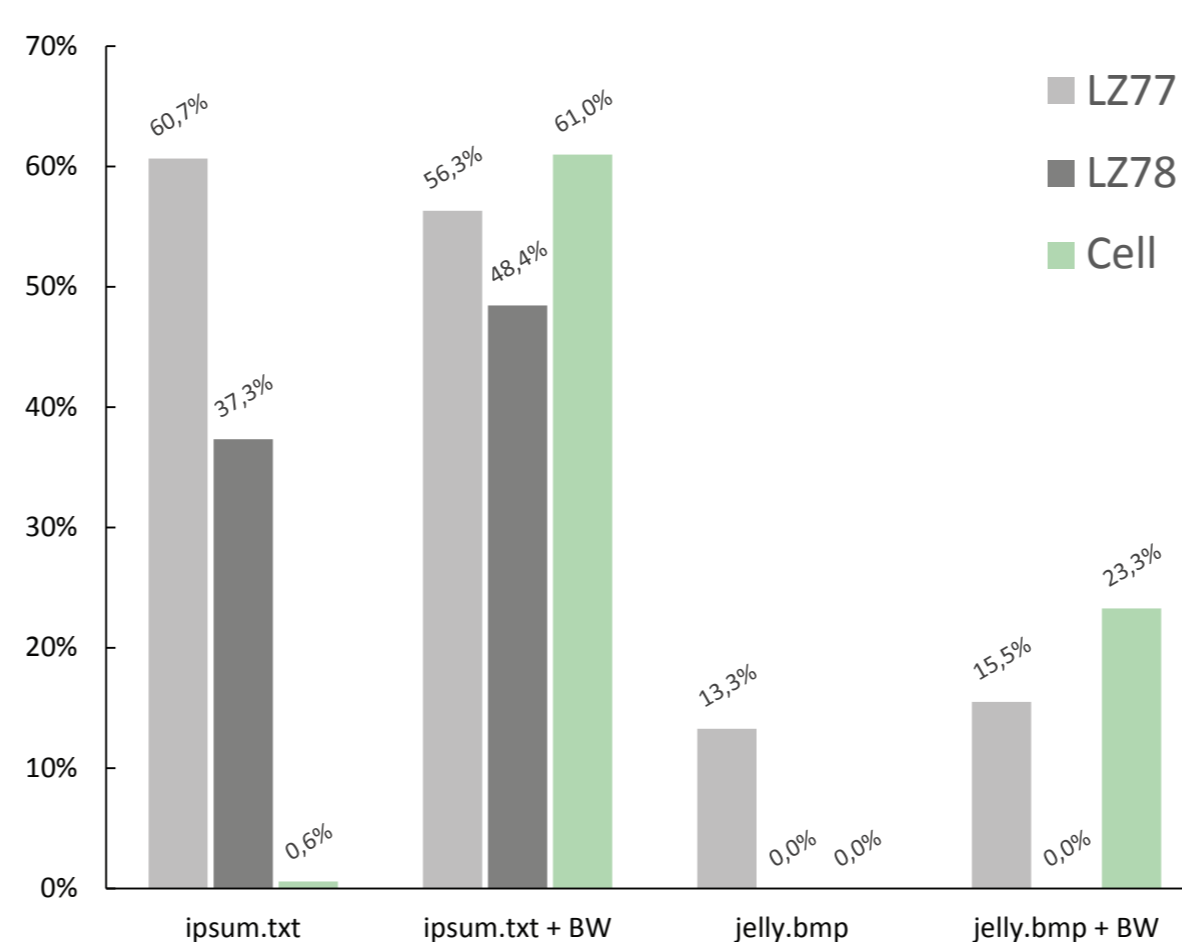


Diagram znázorňuje kompresi textového řetězce „ananas“ pomocí implementace algoritmu s Wolframovým pravidlem 60 a bitovou RLE kompresí. Nejvhodnější stav zde byl nalezen v generaci -51.

Výsledky



Pro transformaci dat bylo zvoleno Wolframovo pravidlo 60 na acyklickém poli buněk. Za kompresní algoritmus byla vybrána RLE komprese operující nad bity místo bajtů či znaků. Počty opakování bitů jsou zapisovány pomocí Gama kódu. Testovacími daty jsou textový soubor ipsum.txt (70,7 kB) a obrazový soubor jelly.bmp (192 kB), pro zlepšení kompresního poměru byla dále na oba vstupní soubory aplikována Burrows-Wheelerova (BW) transformace.

Graf vlevo znázorňuje úsporu místa dosaženou na testovacích souborech a jejich BW-transformovaných variantách pomocí algoritmů LZ77, LZ78 a navrženého algoritmu Cell s maximální testovanou velikostí bloku 4096 bajtů.

*Pojmy

Zpětný běh – hledání stavů předcházejících danému stavu v celulárním automatu. Na rozdíl od stavu následujícího může být přímých předchůdců stavu více či nemusí existovat žádný.

RLE – Run-length encoding. Kódování spočívající v zápisu sekvencí opakujícího se znaku jediným zápisem tohoto znaku spolu s počtem jeho opakování.

Závěr

Vyvinutý algoritmus dosahuje kompresních poměrů srovnatelných se slovníkovými algoritmy LZ77 a LZ78. Pro tuto účinnost musí však být vstupní data nejprve transformována pomocí Burrows-Wheelerovy transformace. Aplikací konvenčních kompresních algoritmů, které kromě slovníkových přístupů implementují často i právě tuto transformaci či jiné metody pro optimalizaci komprese dat, bychom dosáhli lepších výsledků.

Při transformaci dat je navíc prohledáváno a analyzováno velké množství stavů celulárního automatu, časová složitost je tedy vzhledem k dosažené kompresi neúměrná. Výsledek je však postačující k prokázání úvodní hypotézy, že celulární automaty obecně jsou schopny transformovat data způsobem vhodným pro bezztrátovou kompresi.

Literatura

BURROWS M. a WHEELER, D. J. (1994). A Block-sorting Lossless Data Compression Algorithm. Technical report, DIGITAL SRC RESEARCH REPORT.

JERAS, I. a DOBNIKAR, A. (2007). Algorithms for computing preimages of cellular automata configurations. *Physica D: Nonlinear Phenomena*, 233(2), 95-111.

SAYOOD, K. (2012). *Introduction to Data Compression*. Newnes.

WOLFRAM, S. (2002). *A New Kind of Science*. Wolfram media Champaign.