

盤面上の単語配置問題

原口和也 佐藤潤一

小樽商科大学・商学部・社会情報学科

組合せゲーム・パズル プロジェクト・第12回研究集会

2017年3月6日

単語配置問題

($n \times n$ 盤面)

+

ば	な	な		
り	ん	ご		
い	ち	ご		
き	ゆ	う	り	
...				

(辞書)



ふ	き		に	ら
	ゆ			つ
	う	り		き
か	り	ん		よ
き		ご	ぼ	う

(単語の配置)

- 辞書に含まれる単語を盤面上に「うまく」配置することを問う問題
- クロスワードパズル, スケルトンパズルなどの生成問題の一般化

⇒ 整数最適化問題として定式化, ソルバで解いてみた

クロスワードパズルに関する先行研究

- 自動生成 (1980s - 2000s)

- 盤面を彩色し，その上で辞書の中の単語を配置するという枠組み [Smith & Steen, 1981]
- 与えられた彩色に対して解を求める手法
(探索, 制約最適化, ...)

※本研究の定式化では，**彩色も同時に求める**

- 自動解法 (2000s 以降)

- 問題文をNLPで解釈，辞書をIRで構成，解を上記手法で求める
- WebCrow [Ernandes et al., 2005] など

「うまい」配置とは？

- n^2 個のマス目に文字を割り当てる問題と考える
 - 白マス：文字が割り当てられたマス
 - 黒マス：文字が割り当てられないマス
- 目的関数：白マスの数，単語の数，**単語長の総和**， etc.
- 制約条件

- 長さ2以上の極大な白区間には辞書内の単語が現れる
- 同じ単語は高々1回
- 白マスは連結

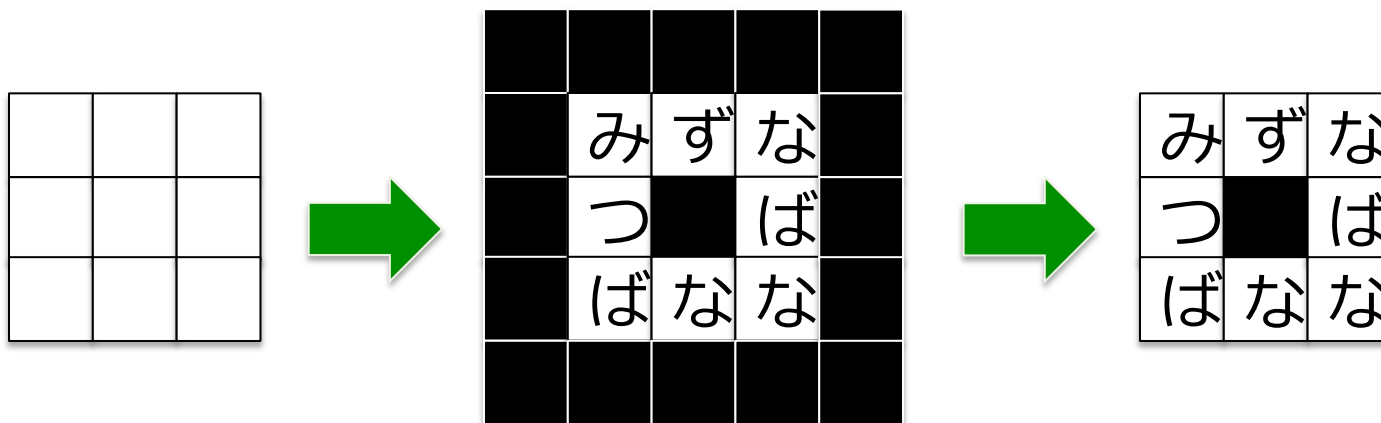
ばなな
りんご
いちご
きゅうり
...

ふ	き	黒	に	ら
黒	ゆ	黒	黒	つ
黒	う	り	黒	き
か	り	ん	黒	よ
き	黒	ご	ぼ	う

※日本のクロスワード：黒マスは隣接せず，四隅は白マス

定式化のアイデア

- 黒マスを表す文字■を文字集合に追加
{あ, い, ..., ん} ∪ {■}
- 辞書内のすべての単語について, その前後に黒マスを追加
ばなな → ■ばなな■
- $(n+2) \times (n+2)$ 盤面上で解を求め, 外周を取り除く



定式化：変数と目的関数

● 主な変数

● 単語の配置 (0-1変数)

- $x[i,j,k] = \begin{cases} 1 & \text{if 単語}k\text{を}(i,j)\text{を先頭に}\underline{\text{横向き}}\text{に配置} \\ 0 & \text{otherwise} \end{cases}$

- 縦向きの配置を表す $y[i,j,k]$ も同様

● 文字の割当 (0-1変数)

- $z[i,j,a] = \begin{cases} 1 & \text{if 文字}a\text{を}(i,j)\text{に割当} \\ 0 & \text{otherwise} \end{cases}$

● フロー (実数変数, 後述)

● 目的関数

- $\sum \text{len}[k] \sum (x[i,j,k] + y[i,j,k]) = \text{単語長の総和}$

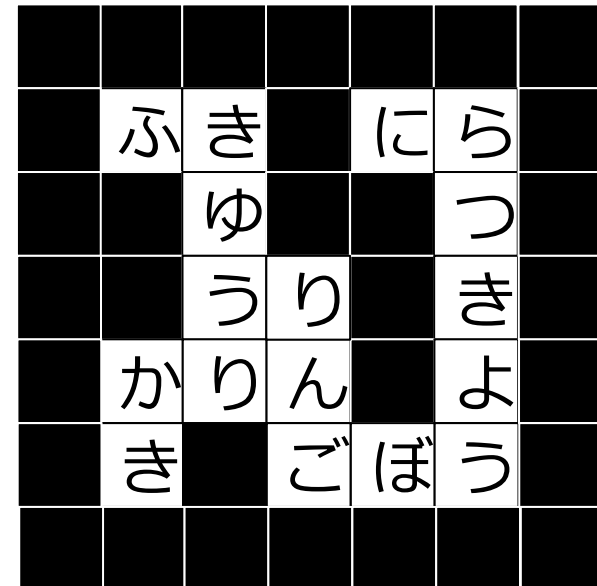
= 白マス数 + 語の交差するマス数 (len[k] は単語 k の長さ)

み	ず	な
つ	■	ば
ば	な	な

(8+4=12)

定式化：制約条件

- 基本的な制約条件
 - マス目には必ず文字が1つ割り当てられる (■含む)
ただし外周には必ず■が割り当てられる
 - マス目には、そこから始まる横単語が高々1つ配置
縦単語も同様
 - 同じ単語は高々1回しか使えない
- 長さ2以上の極大な白区間には
辞書内の単語が現れる
 - 配置した単語と文字が矛盾しない
 - 黒白白 ⇒ 黒を起点とした単語を配置

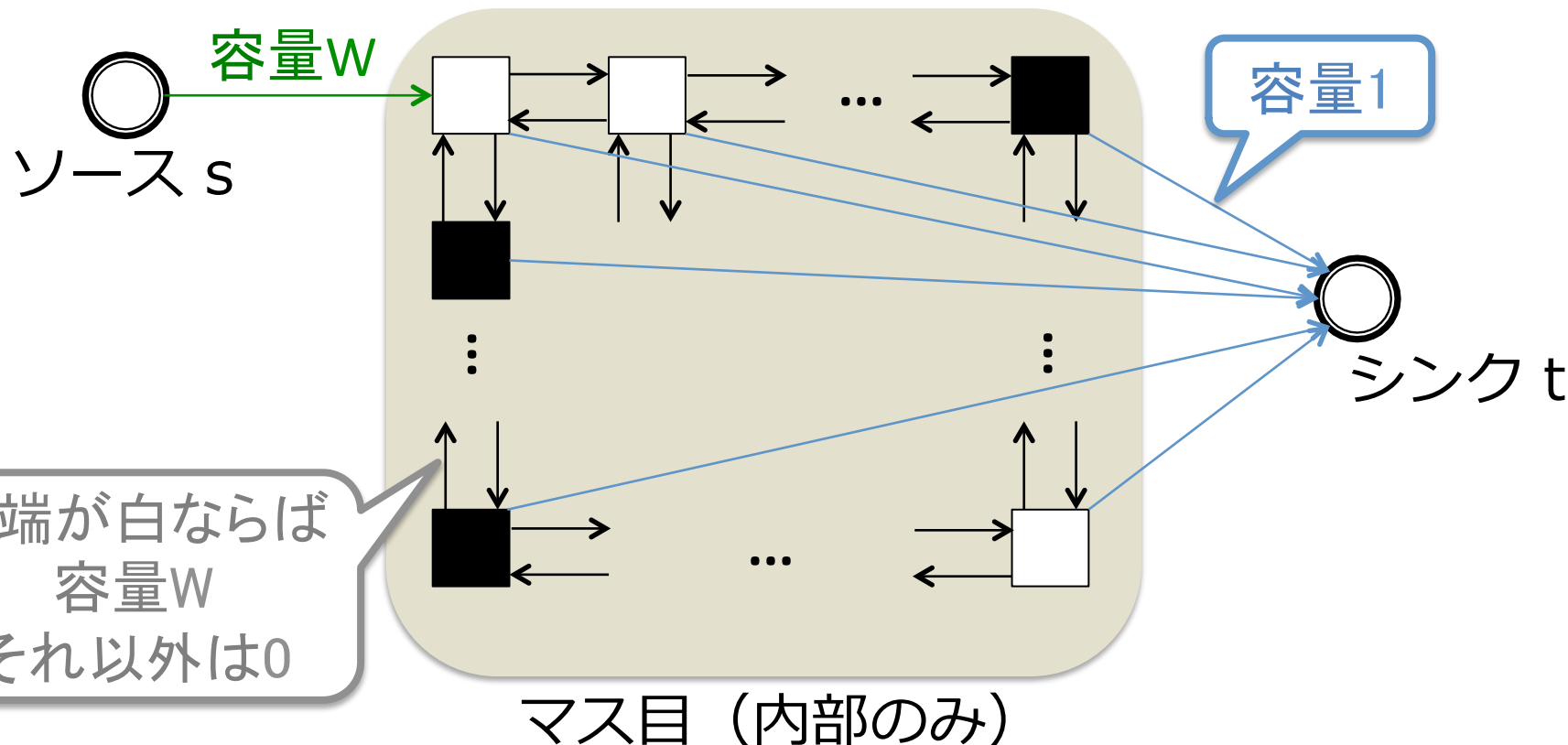


定式化：制約条件

- 基本的な制約条件
 - マス目には必ず文字が1つ割り当てられる (■含む)
ただし外周には必ず■が割り当てられる
 - マス目には, そこから始まる横単語が高々1つ配置
縦単語も同様
 - 同じ単語は高々1回しか使えない
- 長さ2以上の極大な白区間には
辞書内の単語が現れる
 - 配置した単語と文字が矛盾しない
 - 黒白白 ⇒ 黒を起点とした単語を配置



- 白マスは連結 = 以下の(s,t)-ネットワークにおいて、フロー量 W のフローが存在 (W は白マスの総数)



- フロー量 W のフローは、すべての白マスを通り、各白マスからシンクに1ずつ流れる
- フローを表す変数は連続変数で取り扱うことができる

問題のサイズ

- 変数の数 $O((m+q)n^2)$
 - m は辞書のサイズ
 - q は文字の個数
- 制約式の数 $O(mn^3)$

既成のパズルの再現

～解から辞書を構成し，元のパズルを再現できるか～

- ネットで拾った日本語クロスワードパズル

	n=6	7	8	9	10
成功率	3/3	3/3	2/3	0/3	0/3
変数 ($\times 10^3$)	1.9	2.9	5.1	7.9	10.4
制約 ($\times 10^3$)	4.7	8.3	14.5	24.6	31.4

※ 計算環境

- ソルバ：IBM ILOG CPLEX v12.6.1
- 計算機：CPUのクロック周波数は3.40GHz, メインメモリの容量は8GB
- 計算時間600秒で打ち切り

特徴的な辞書からスケルトンを生成

言語	辞書	単語数	n=6	7	8	9	10
Ja	小樽	40	63.9	63.3	53.1	56.8	19.0
	五輪	60	69.4	65.3	34.4	8.6	6.0
	漫画	80	66.7	63.3	56.3	11.1	2.0
	野球	100	66.7	63.3	15.6	4.9	6.0
	北海道	120	66.7	38.8	7.8	12.3	0.0
	野菜	140	66.7	16.3	7.8	0.0	0.0
En	世界	206	72.2	73.5	64.1	18.5	0.0

数値は得られた解における白マスの割合 (%), 赤は最適解

た	か	ひ	ろ		に	し
	く					ゆ
	な	か	た	し	よ	う
	か		け			い
	か		や	ま	ぐ	ち
う	つ	み				
	や	ま	だ	て	つ	と

T	U	N	I	S	I	A
U	S	A		U		L
V		M		D		B
A		I		A		A
L	E	B	A	N	O	N
U		I				I
	P	A	N	A	M	A

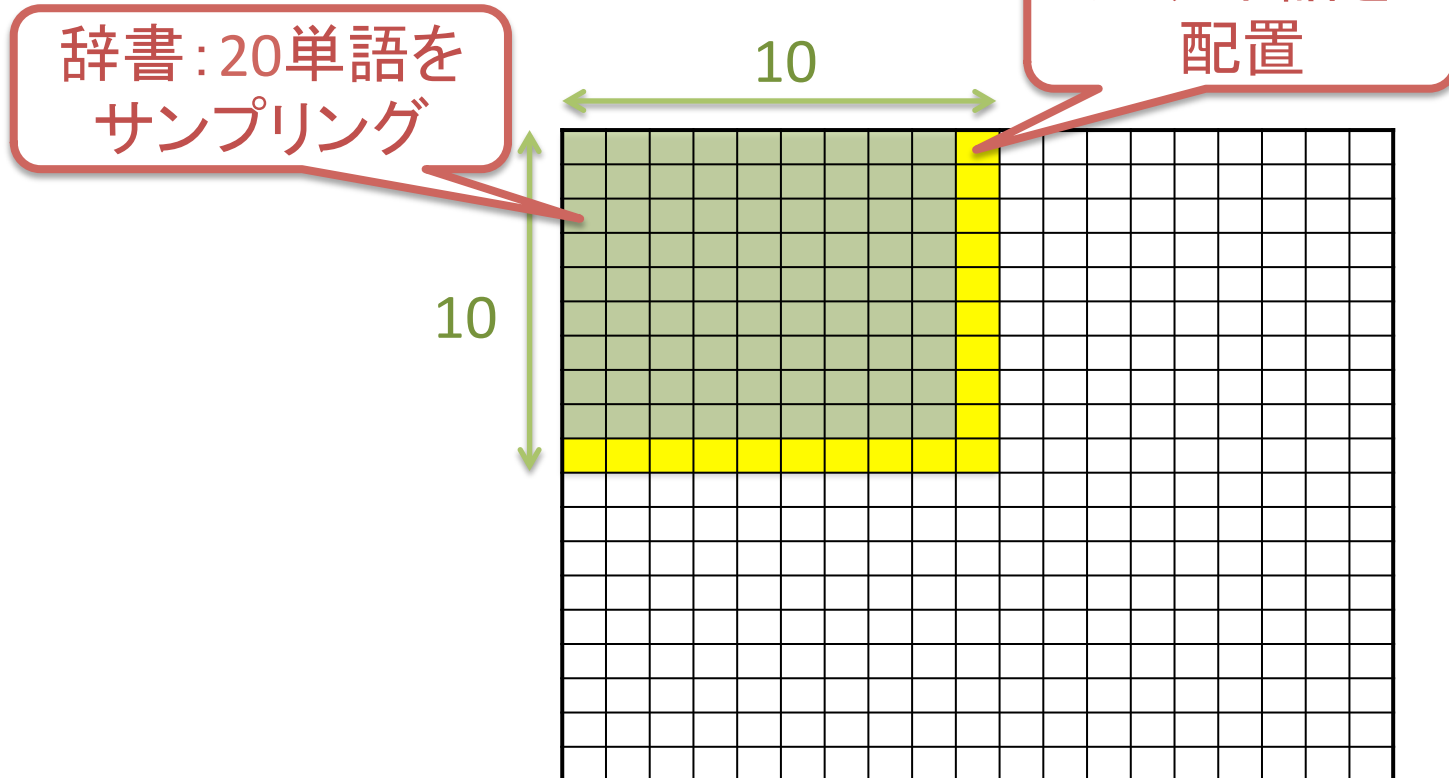
辞書: 野球
 白マスの割合: 63.3%
 語数/n: 1.43

辞書: 世界
 白マスの割合: 73.5%
 語数/n: 1.29

使われる単語数はおおむね2n以下

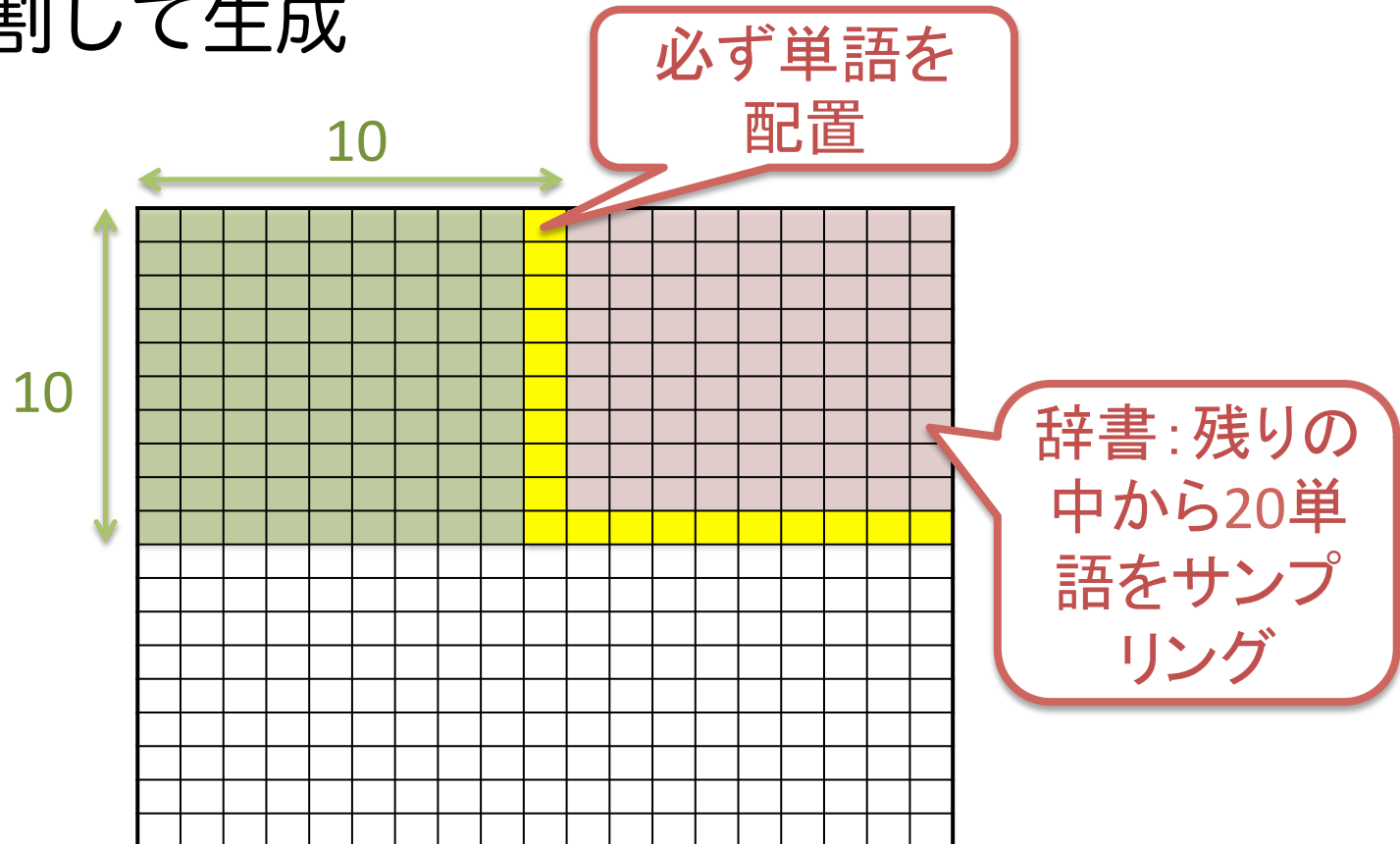
次数nが大きいスケルトンの生成

- ニコリ2017冬号のスケルトンコンテスト
 - n=19, 60単語（帆船名），単語には重み（帆の面積）
 - 重みの総和が最大となるように単語を配置せよ
- 方針：分割して生成



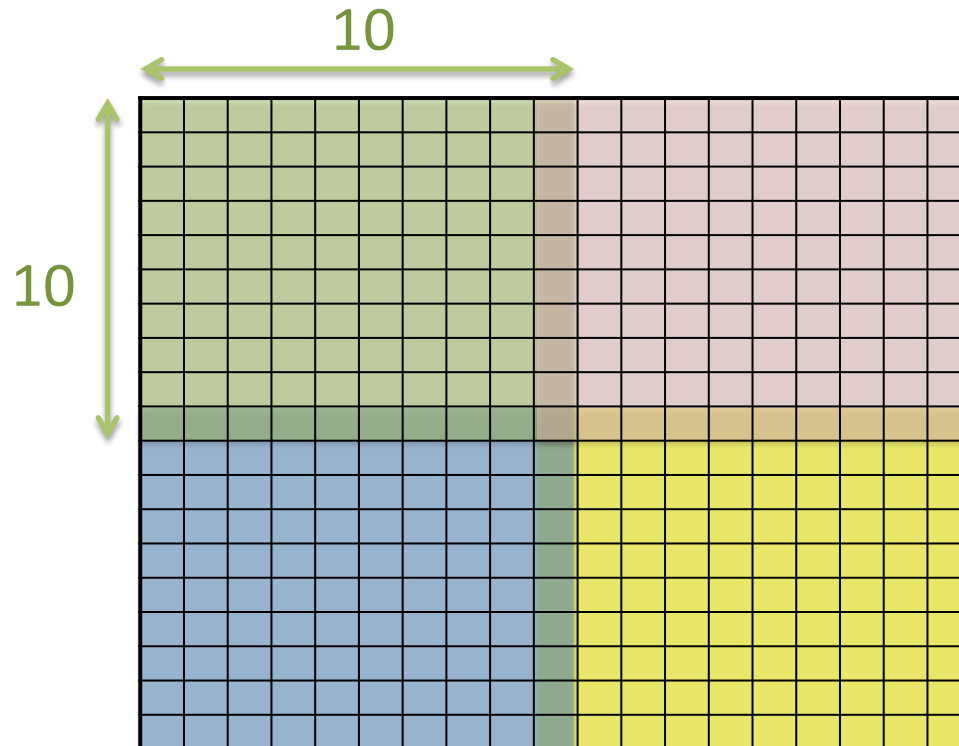
次数nが大きいスケルトンの生成

- ニコリ2017冬号のスケルトンコンテスト
 - n=19, 60単語（帆船名），単語には重み（帆の面積）
 - 重みの総和が最大となるように単語を配置せよ
- 方針：分割して生成



次数nが大きいスケルトンの生成

- ニコリ2017冬号のスケルトンコンテスト
 - n=19, 60単語（帆船名），単語には重み（帆の面積）
 - 重みの総和が最大となるように単語を配置せよ
- 方針：分割して生成



							G		S	O	R	L	A	N	D	E	T			
M		E		E			L		E				T					P		
I		S	E	A	C	L	O	U	D		P	A	L	I	N	U	R	O		
R		M		G			R		O				A					L		
C	R	E	O	L	E		I		V				N					Y		
E		R		E			A			F	A	N	T	O	M	E		N		
A		A									D		I					E		
		L	I	B	E	R	T	A	D		I		S	A	G	R	E	S		
		D									X							I		
S	T	A	R	F	L	Y	E	R				P	O	G	O	R	I	A		
					O						Z		U							
H	M	S	W	A	R	R	I	O	R		E		T							
					D							B	E	L	E	M		G		
N	I	P	P	O	N	M	A	R	U		U		A		O			U		
					E								W		S			A		
					L										H			Y		
C	U	T	T	Y	S	A	R	K			B	A	L	C	L	U	T	H	A	
					O											L			S	
					D	A	N	M	A	R	K			M	A	R	Q	U	E	S

上の解の重みは**58388**

(良いかどうかは次号以降をお楽しみに)

まとめ

- 単語配置問題を，混合0-1整数線形最適化問題として定式化し，いくつかの実験を行った
 - 辞書にもよるが， $n \leq 10$ 程度なら短時間で生成できる
 - n が大きい場合も，切り貼りによって生成できる
- メモ
 - 局所探索などでさらに改善することも可能
 - 作家さんは偉大