# Génération des multiensembles fréquents de sous-mots partitionnant un mot

Julien David     Lhouari Nourine

LIMOS - Université Blaise Pascal
ANR DAG (ANR-09-EMER-003-01)

1er avril 2010

# Définition du problème

- **Données :** un alphabet $\Sigma$, un mot $w \in \Sigma^+$, un entier $q$.
- **Problème :** Engendrer toutes les partitions de $w$ en sous-mots fréquents.

**Exemple**

*Soit le mot $w = aaabba$ et un entier $q = 2$.*
*Les sous-mots a et ab permettent de partitionner w et chaque sous-mot apparait au moins q fois la partition.*

# Motivations

- Algorithmique sur les mots.
- Log mining :
  - Restitution des corrélations dans une base de données,
  - Analyse de systèmes multiprocesseurs.

# Plan

1. Produit de mélange, multiensembles de mots
2. Graphe des multiensembles
3. Algorithme et complexité

# Ordre militaire

L'ordre militaire, ou ordre lexicographique gradué, sur les mots est défini comme suit :

$$\forall v, w \in \Sigma^*,\ v <_{mil} w \iff \begin{cases} |v| < |w| \\ \text{ou} \\ |v| = |w| \text{ et } v <_{lex} w \end{cases}$$

# Produit de mélange

Le produit de mélange de deux mots $u$ et $v$ est l'ensemble :

$$u \sqcup\!\sqcup v = \{u_1 v_1 \dots u_n v_n \mid u_i, v_i \in \Sigma^* \text{ pour } 1 \leq i \leq n, \\ u = u_1 \dots u_n \text{ et } v = v_1 \dots v_n\}.$$

**Exemple**

$$ab \sqcup\!\sqcup bc = \{abbc, abcb, babc, bacb, bcab\}$$

# Produit de mélange

Le produit de mélange de deux mots $u$ et $v$ est l'ensemble :

$$u \sqcup\!\sqcup v = \{u_1 v_1 \ldots u_n v_n \mid u_i, v_i \in \Sigma^* \text{ pour } 1 \leq i \leq n, \\ u = u_1 \ldots u_n \text{ et } v = v_1 \ldots v_n\}.$$

Il est possible d'étendre cet opérateur :

$$\sqcup\!\sqcup^{k+1} w = (\sqcup\!\sqcup^{k} w) \sqcup\!\sqcup w$$

$$\sqcup\!\sqcup^{0} w = \varepsilon$$

Soit un ensemble de mots $X = \{w_1, \ldots, w_n\}$

$$\bigsqcup\!\!\bigsqcup X = w_1 \sqcup\!\sqcup w_2 \sqcup\!\sqcup \ldots \sqcup\!\sqcup w_n$$

# Sous-mots

Un mot $v$ est un sous-mot d'un mot $w$ si et seulement si

$$\text{il existe un mot } u \in \Sigma^* \text{ tel que } w \in v \sqcup\!\sqcup u.$$

Un ensemble $X$ de sous-mots permet de partitionner un mot $w$ si et seulement si

$$w \in \bigsqcup\!\!\bigsqcup X$$

# Multiensembles de mots

Un multiensemble de mots $\mathcal{M}$ est un couple $\langle X, f\rangle$ tel que

- $X \in 2^{\Sigma^+}$ est l'ensemble de mots sous-jacent,
- $f : \Sigma^+ \mapsto \mathbb{N}$ associe à chaque mot sa multiplicité dans $\mathcal{M}$.

Soit un multiensemble $\mathcal{M} = \langle X, f\rangle$. On étend le produit de mélange aux multiensembles

$$\bigsqcup\!\bigsqcup \mathcal{M} = \bigsqcup\!\bigsqcup_{v \in X} (\sqcup\!\sqcup^{f(v)} v).$$

# Multiensembles de mots

Un multiensemble $\mathcal{M}$ de sous-mots permet de partitionner un mot $w$ si et seulement si

$$w \in \bigsqcup \mathcal{M}$$

**Exemple**

*Soit le mot* $w = bbabccbcab$ *et un entier* $q = 2$*, le multiensemble*

$$\{(ab, 2), (bc, 3)\}$$

*est une solution. En effet,*

$$abab \in \sqcup\!\sqcup^2 \ ab$$

$$bbccbc \in \sqcup\!\sqcup^3 \ bc$$

$$bbabccbcab \in abab \sqcup\!\sqcup bbccbc$$

# Multiensembles candidats

**Definition**

Soient $\mathcal{M} = \langle X, f \rangle$ un multiensemble, $w \in \Sigma^+$ un mot et $q \in \mathbb{N}$ un entier.
On dit que $\mathcal{M}$ est un $(w, q)$-candidat si :

1. $\forall v \in \Sigma^+$, soit $f(v) \geq q$ soit $f(v) = 0$,
2. pour tout $a \in \Sigma$,
$$\sum_{v \in X} (f(v) \times \|v\|_a) = \|w\|_a.$$

On note $\mathfrak{M}_{w,q}$ l'ensemble des $(w, q)$-candidats.

# Multiensembles valides

## Definition

Un multiensemble $\mathcal{M} = \langle X, f \rangle$ est $(w, q)$-valide si :

1. $\mathcal{M}$ est un $(w, q)$-candidat,
2. Le mot $w$ appartient au produit de mélange de $\mathcal{M}$

$$w \in \bigsqcup\!\!\!\bigsqcup \mathcal{M}$$

On note $\mathcal{M}_{w,q}$ l'ensemble des multiensembles $(w, q)$-valides.
On a

$$\mathcal{M}_{w,q} \subseteq \mathfrak{M}_{w,q}.$$

# Génération des solutions

**Méthode**

Afin d'engendrer les multiensembles $(w, q)$-valides, on utilise la méthode suivante :

- on engendre exhaustivement et sans redondance les $(w, q)$-candidats,
- pour chaque candidat, on teste s'il s'agit d'un multiensemble $(w, q)$-valide.

On commence par présenter un ensemble d'opérateurs sur les multiensembles.

# Sommes de multiensembles

## Definition

Soient $\mathcal{M}_1 = \langle X, f \rangle$ et $\mathcal{M}_2 = \langle Y, g \rangle$ deux multiensembles.
La somme de $\mathcal{M}_1$ et $\mathcal{M}_2$, notée $\mathcal{M}_1 \oplus \mathcal{M}_2$, est le multiensemble $\mathcal{M} = \langle Z, h \rangle$ défini comme suit :

- $Z = X \cup Y$,
- pour tout $z \in Z$, on a $h(z) = f(z) + g(z)$.

## Exemple

$$\{(aa, 1), (ba, 4)\} \oplus \{(aa, 2), (bac, 3)\} = \{(aa, 3), (ba, 4), (bac, 3)\}$$

# Différence de multiensembles

## Definition

Soient $\mathcal{M}_1 = \langle X, f \rangle$ et $\mathcal{M}_2 = \langle Y, g \rangle$ deux multiensembles.
La différences de $\mathcal{M}_1$ et $\mathcal{M}_2$, notée $\mathcal{M}_1 \ominus \mathcal{M}_2$, est le multiensemble $\mathcal{M} = \langle Z, h \rangle$ défini comme suit :

- pour tout $x \in X$, on a $h(z) = max\{f(z) - g(z), 0\}$.

## Exemple

$$\{(aa, 1), (ba, 4)\} \ominus \{(aa, 2), (ba, 3)\} = \{(ba, 1)\}$$

# Extension de multiensemble

**Definition**

Soient un multiensemble $\mathcal{M} = \langle X, f \rangle$ un multiensemble un mot $v \in X$, une lettre $a \in X \cap \Sigma$ et un entier positif $i$. On définit le multiensemble $\mathfrak{e}(\mathcal{M}, v, a, i)$ obtenu en appliquant les opérations suivantes :

$$\mathfrak{e}(\mathcal{M}, v, a, i) = \mathcal{M} \oplus \{(va, i)\} \ominus \{(v, i)\} \ominus \{(a, i)\}$$

**Exemple**

Soit le multiensemble $\mathcal{M} = \{(a, 1), (ba, 4)\}$, on a

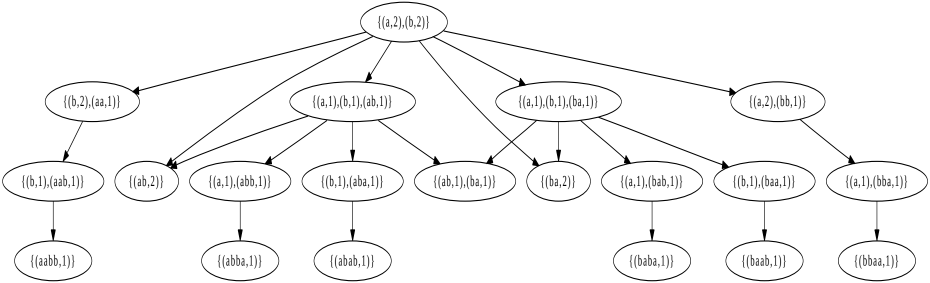
$$\mathfrak{e}(\mathcal{M}, ba, a, 1) = \{(ba, 3), (baa, 1)\}$$

# Le graphe de transition

Le graphe de transition $T_{w,q} = (\mathfrak{M}_{w,q}, E_{w,q})$ est un graphe orienté défini comme suit :

- son ensemble de sommets $\mathfrak{M}_{w,q}$ est l'ensemble des $(w, q)$-candidats,
- Il existe une arête $(\mathcal{M}_1, \mathcal{M}_2) \in E_{w,q}$ s'il existe un entier $i$, un mot $v$ et une lettre $a$ tels que $\mathcal{M}_2 = \mathfrak{e}(\mathcal{M}_1, v, a, i)$.

# Le graphe de transition $T_{abab,1}$



# Propriétés du graphe de transitions

Pour tout graphe $T_{w,q}$, les propriétés suivantes sont satisfaites :

- le graphe de transition est acyclique et sa hauteur est bornée par $|w|$.
- le graphe de transition admet un arbre couvrant.
- L'ensemble des multiensembles $(w, q)$-valides forme un sous-arbre de l'arbre couvrant.

# Graphe acyclique

**Lemma**

*Le graphe $T_{w,q}$ est acyclique et sa hauteur est bornée par $|w|$.*

**Preuve**

- Il existe un arc $(\mathcal{M}_1, \mathcal{M}_2) \in E_{w,q}$ si et seulement si il existe un mot $v$, une lettre $a$ et un entier $i$ tel que $\mathcal{M}_2 = \mathfrak{e}(\mathcal{M}_1, v, a, i)$.
- Le nombre d'occurrence de la lettre $a$ dans $\mathcal{M}_2$ est strictement inférieur au nombre d'occurrence de $a$ dans $\mathcal{M}_1$.
- S'il existe un chemin entre deux multiensembles dans $T_{w,q}$, il existe une lettre $a \in \Sigma$, pour laquelle les nombres d'occurrences diffèrent d'un multiensemble à l'autre.
- La somme du nombre d'occurences de chaque lettre dans un $(w, q)$-candidat est au plus égal à $|w|$. La longueur de plus long chemin dans $T_{w,q}$ est donc bornée par $|w|$.

# Graphe acyclique

## Lemma

*Le graphe $T_{w,q}$ est acyclique et sa hauteur est bornée par $|w|$.*

## Preuve

- Il existe un arc $(\mathcal{M}_1, \mathcal{M}_2) \in E_{w,q}$ si et seulement si il existe un mot $v$, une lettre $a$ et un entier $i$ tel que $\mathcal{M}_2 = \mathfrak{e}(\mathcal{M}_1, v, a, i)$.
- Le nombre d'occurrence de la lettre $a$ dans $\mathcal{M}_2$ est strictement inférieur au nombre d'occurrence de $a$ dans $\mathcal{M}_1$.
- S'il existe un chemin entre deux multiensembles dans $T_{w,q}$, il existe une lettre $a \in \Sigma$, pour laquelle les nombres d'occurrences diffèrent d'un multiensemble à l'autre.
- La somme du nombre d'occurences de chaque lettre dans un $(w, q)$-candidat est au plus égal à $|w|$. La longueur de plus long chemin dans $T_{w,q}$ est donc bornée par $|w|$.

# Solution Triviale

- On définit le multiensemble $\mathcal{M}^0_{w,q}$ :
  - Son ensemble de mot sous-jacent est l'alphabet $\Sigma$ sur lequel $w$ est défini.
  - Le nombre d'occurrences de chaque lettre $a$ dans $\mathcal{M}^0_{w,q}$ est égal à $\|w\|_a$.
- Pour tout $q \leq \min\{\|w\|_a \mid a \in \Sigma\}$, $\mathcal{M}^0_{w,q}$ est $(w,q)$-valide car

$$w \in \bigsqcup_{a\in\Sigma} (\sqcup^{\|w\|_a}\ a).$$

- $\mathcal{M}^0_{w,q}$ est l'unique $(w,q)$-candidat dont l'ensemble de mots associé ne contient que des lettres.

# L'ensemble des arêtes $E_{w,q}$

Soit un multiensemble $\mathcal{M} = \langle X, f\rangle$ qui est $(w,q)$-candidat, un mot $v \in X$ et une lettre $a \in X \cap \Sigma$.
Pour tout entier $i \in \mathbb{N}*$ tel que $\mathfrak{e}(\mathcal{M}, v, a, i) \in \mathfrak{M}_{w,q}$, on a :

- Si $v \neq a$, alors
  - $f(va) + i \geq q$,
  - $f(v) - i \geq q$ ou $f(v) - i = 0$,
  - $f(a) - i \geq q$ ou $f(a) - i = 0$.
- Si $v = a$, alors
  - $f(aa) + i \geq q$,
  - $f(a) - 2i \geq q$ ou $f(a) - 2i = 0$.

# L'ensemble des arêtes $E_{w,q}$

- On note $\mathcal{I}(\mathcal{M}, v, a)$ l'ensemble des entiers $i$ tel que $\mathfrak{e}(\mathcal{M}, v, a, i) \in \mathfrak{M}_{w,q}$.
- Il est possible de déterminer si $\mathcal{I}(\mathcal{M}, v, a) = \emptyset$ ou de calculer $\min(\mathcal{I}(\mathcal{M}, v, a))$ en temps constant.

**Exemple**

- Si $v \neq a$, $f(va) \geq q$, $f(v) > q$ $f(a) > q$, alors $\min(\mathcal{I}(\mathcal{M}, v, a)) = 1$.
- Si $v \neq a$, $f(va) = 0$, $f(v) \geq 2q$ $f(a) \geq 2q$, alors $\min(\mathcal{I}(\mathcal{M}, v, a)) = q$.
- ...

# Arbre couvrant

L'arbre couvrant est un sous-ensemble d'arêtes $(\langle X, f\rangle, \langle Y, g\rangle)$ telles qu'il existe un mot $v \in X$, une lettre $a \in X$ et un entier $i$ tel que :

- $\langle Y, g\rangle = e(X, f, v, a, i)$,
- $va = max_{mil}(Y)$,
- $i = min(\mathcal{I}(X, f, q, v, a))$.

L'arbre couvrant est enraciné en $\mathcal{M}^0_{w,q}$ et sa hauteur est borné par $|w|$.

# Relation antimonotone

**Lemme**

*Pour tous multiensembles $\mathcal{M}_1, \mathcal{M}_2 \in \mathfrak{M}_{w,q}$ tels que $(\mathcal{M}_1, \mathcal{M}_2) \in E_{w,q}$, si $\mathcal{M}_2$ est $(w,q)$-valide alors $\mathcal{M}_1$ est aussi $(w,q)$-valide.*

$\mathcal{M}_1, \mathcal{M}_2$ étant des $(w,q)$-candidats, on montre que

$$w \in \bigsqcup \mathcal{M}_2 \Longrightarrow w \in \bigsqcup \mathcal{M}_1.$$

# Relation antimonotone

Pour tout mot $v$ et toute lettre $a$, on a

$$va \subset v \sqcup\!\sqcup a$$

Pour tout multiensemble $\mathcal{M} = \langle X, f \rangle$, on a

$$\left(\bigsqcup\!\bigsqcup \mathcal{M} \sqcup\!\sqcup va\right) \subset \left(\bigsqcup\!\bigsqcup \mathcal{M} \sqcup\!\sqcup v \sqcup\!\sqcup a\right).$$

Et par induction,

$$\left(\bigsqcup\!\bigsqcup \mathcal{M} \sqcup\!\sqcup^i va\right) \subset \left(\bigsqcup\!\bigsqcup \mathcal{M} \sqcup\!\sqcup^i v \sqcup\!\sqcup^i a\right),$$

$$\bigsqcup\!\bigsqcup (\mathcal{M} \oplus \{(va, i)\}) \subset \bigsqcup\!\bigsqcup (\mathcal{M} \oplus \{(v, i)\} \oplus \{(a, i)\})$$

Pour tout entier $i$ tel que $i \leq f(v)$ et $i \leq f(a)$ on a bien

$$\left(\bigsqcup\!\bigsqcup \mathcal{M} \oplus \{(va, i)\} \ominus \{(v, i)\} \ominus \{(a, i)\}\right) \subset \bigsqcup\!\bigsqcup \mathcal{M}$$

$$w \in \bigsqcup\!\bigsqcup \mathfrak{e}(\mathcal{M}, v, a, i) \Longrightarrow w \in \bigsqcup\!\bigsqcup \mathcal{M}$$

# Relation antimonotone

Pour tout mot $v$ et toute lettre $a$, on a

$$va \subset v \sqcup\!\sqcup a$$

Pour tout multiensemble $\mathcal{M} = \langle X, f \rangle$, on a

$$\left(\bigsqcup\!\bigsqcup \mathcal{M} \sqcup\!\sqcup va\right) \subset \left(\bigsqcup\!\bigsqcup \mathcal{M} \sqcup\!\sqcup v \sqcup\!\sqcup a\right).$$

Et par induction,

$$\left(\bigsqcup\!\bigsqcup \mathcal{M} \sqcup\!\sqcup^i va\right) \subset \left(\bigsqcup\!\bigsqcup \mathcal{M} \sqcup\!\sqcup^i v \sqcup\!\sqcup^i a\right),$$

$$\bigsqcup\!\bigsqcup (\mathcal{M} \oplus \{(va, i)\}) \subset \bigsqcup\!\bigsqcup (\mathcal{M} \oplus \{(v, i)\} \oplus \{(a, i)\})$$

Pour tout entier $i$ tel que $i \leq f(v)$ et $i \leq f(a)$ on a bien

$$\left(\bigsqcup\!\bigsqcup \mathcal{M} \oplus \{(va, i)\} \ominus \{(v, i)\} \ominus \{(a, i)\}\right) \subset \bigsqcup\!\bigsqcup \mathcal{M}$$

$$w \in \bigsqcup\!\bigsqcup \mathfrak{e}(\mathcal{M}, v, a, i) \Longrightarrow w \in \bigsqcup\!\bigsqcup \mathcal{M}$$

# Relation antimonotone

Pour tout mot $v$ et toute lettre $a$, on a

$$va \subset v \sqcup\!\sqcup a$$

Pour tout multiensemble $\mathcal{M} = \langle X, f \rangle$, on a

$$\left( \bigsqcup\!\!\bigsqcup \mathcal{M} \sqcup\!\sqcup va \right) \subset \left( \bigsqcup\!\!\bigsqcup \mathcal{M} \sqcup\!\sqcup v \sqcup\!\sqcup a \right).$$

Et par induction,

$$\left( \bigsqcup\!\!\bigsqcup \mathcal{M} \sqcup\!\sqcup^{i} va \right) \subset \left( \bigsqcup\!\!\bigsqcup \mathcal{M} \sqcup\!\sqcup^{i} v \sqcup\!\sqcup^{i} a \right),$$

$$\bigsqcup\!\!\bigsqcup (\mathcal{M} \oplus \{(va, i)\}) \subset \bigsqcup\!\!\bigsqcup (\mathcal{M} \oplus \{(v, i)\} \oplus \{(a, i)\})$$

Pour tout entier $i$ tel que $i \leq f(v)$ et $i \leq f(a)$ on a bien

$$\left( \bigsqcup\!\!\bigsqcup \mathcal{M} \oplus \{(va, i)\} \ominus \{(v, i)\} \ominus \{(a, i)\} \right) \subset \bigsqcup\!\!\bigsqcup \mathcal{M}$$

$$w \in \bigsqcup\!\!\bigsqcup \mathfrak{e}(\mathcal{M}, v, a, i) \Longrightarrow w \in \bigsqcup\!\!\bigsqcup \mathcal{M}$$

# Relation antimonotone

Pour tout mot $v$ et toute lettre $a$, on a

$$va \subset v \sqcup\!\sqcup a$$

Pour tout multiensemble $\mathcal{M} = \langle X, f \rangle$, on a

$$\left(\bigsqcup\!\bigsqcup \mathcal{M} \sqcup\!\sqcup va\right) \subset \left(\bigsqcup\!\bigsqcup \mathcal{M} \sqcup\!\sqcup v \sqcup\!\sqcup a\right).$$

Et par induction,

$$\left(\bigsqcup\!\bigsqcup \mathcal{M} \sqcup\!\sqcup^{i} va\right) \subset \left(\bigsqcup\!\bigsqcup \mathcal{M} \sqcup\!\sqcup^{i} v \sqcup\!\sqcup^{i} a\right),$$

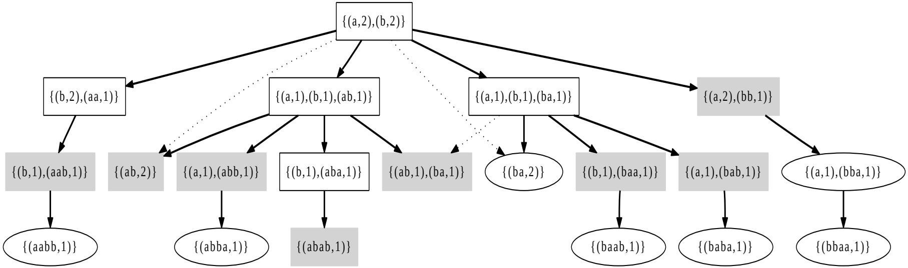
$$\bigsqcup\!\bigsqcup (\mathcal{M} \oplus \{(va, i)\}) \subset \bigsqcup\!\bigsqcup (\mathcal{M} \oplus \{(v, i)\} \oplus \{(a, i)\})$$

Pour tout entier $i$ tel que $i \leq f(v)$ et $i \leq f(a)$ on a bien

$$\left(\bigsqcup\!\bigsqcup \mathcal{M} \oplus \{(va, i)\} \ominus \{(v, i)\} \ominus \{(a, i)\}\right) \subset \bigsqcup\!\bigsqcup \mathcal{M}$$

$$w \in \bigsqcup\!\bigsqcup \mathfrak{e}(\mathcal{M}, v, a, i) \Longrightarrow w \in \bigsqcup\!\bigsqcup \mathcal{M}$$

# L'arbre couvrant de $T_{abab,1}$



# Description de l'algorithme

**Algorithm 1**: Generateur($\mathcal{M}$, $q$, $w$)

**Données**: Un multiensemble $\mathcal{M} = \langle X, f \rangle$, un entier $q$, un mot $w$.

**début**

    **si** $\mathcal{M}$ *est* $(w, q)$*-valide* **alors**
        Afficher $\mathcal{M}$
        **pour tous les** $v \in X$ **faire**
            **pour tous les** $a \in X \cap \Sigma$ **faire**
                **si** $va = max_{mil}\{X \cup va\}$ **alors**
                    **si** $\mathcal{I}(\mathcal{M}, v, a) \neq \emptyset$ **alors**
                        $i = min(\mathcal{I}(\mathcal{M}, v, a))$
                        $\mathcal{M}' = \mathcal{M} \oplus (va, i) \ominus (v, i) \ominus (a, i)$
                        Generateur($\mathcal{M}'$, $q$, $w$)
                    **fin**
                **fin**
            **fin**
        **fin**
    **fin**

**fin**

# Remarque sur la complexité de l'algorithme

La complexité de l'algorithme *Generateur* dépends de

- Du nombre de multiensembles $(w, q)$-valides
  $\rightarrow$ on mesure la complexité en fonction de la taille de la sortie.
- De la complexité de l'oracle qui permet de tester si un multiensemble est $(w, q)$-valide.

# Problème : *Frequent Partition Generation*

- **Données :**
  - un alphabet $\Sigma$,
  - un mot $w \in \Sigma^+$,
  - un entier $q$,
- **Question :** Quel est l'ensemble des multiensembles $(w, q)$-valides ?

# Problème : *Shuffle Product Testing*

- **Données :**
  - un alphabet $\Sigma$,
  - un mot $w \in \Sigma^+$,
  - un entier $q$,
  - un multiensemble $\mathcal{M}$,
- **Question :** $\mathcal{M}$ est-il $(w, q)$-valide ?

# Problème : *Shuffle Product Testing*

**Théorème[Warmuth,Haussler]**

Soit $X$ un ensemble de mots et $w$ un mot. Le problème de décider si $w \in \bigsqcup X$ est NP-complet.

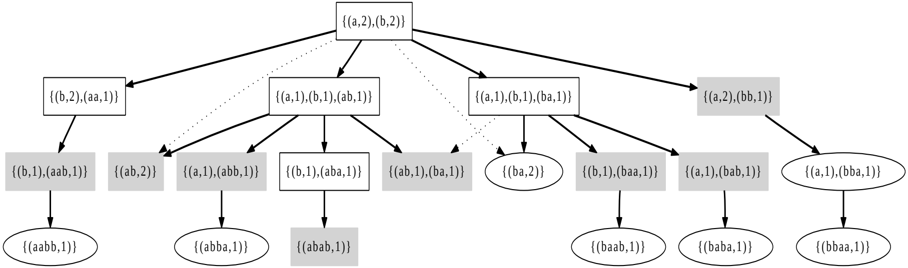
# Problème : *Incremental Shuffle Product Testing*

- **Données :**
  - un alphabet $\Sigma$,
  - un mot $w \in \Sigma^+$,
  - un entier $q$,
  - un multiensemble $(w, q)$-valide $\mathcal{M}$,
  - un multiensemble $\mathcal{M}'$ tel que $\mathcal{M}' = \mathfrak{e}(\mathcal{M}, v, a)$.
- **Question :** $\mathcal{M}'$ est-il $(w, q)$-valide ?

# Problème : *Incremental Shuffle Product Testing*

**Théorème**

Sauf si P=NP, il n'existe pas d'algorithme polynomial pour résoudre *Incremental Shuffle Product Testing*.


{(a,2),(b,2)}
{(b,2),(aa,1)}
{(a,1),(b,1),(ab,1)}
{(a,1),(b,1),(ba,1)}
{(a,2),(bb,1)}
{(b,1),(aab,1)}
{(ab,2)}
{(a,1),(abb,1)}
{(b,1),(aba,1)}
{(ab,1),(ba,1)}
{(ba,2)}
{(b,1),(baa,1)}
{(a,1),(bab,1)}
{(a,1),(bba,1)}
{(aabb,1)}
{(abba,1)}
{(abab,1)}
{(baab,1)}
{(baba,1)}
{(bbaa,1)}

# Réduire le temps de calcul

- On fixe un entier $k$ tel que pour toute solution $\langle X, f \rangle$, on ait

$$\sum_{v \in X} f(v) \leq k$$

$\rightarrow$ tester si $w \in \coprod \langle X, f \rangle$ est réalisable en $\mathcal{O}(|w|^k)$.

- Ajouter des contraintes supplémentaires

# X-factorisation d'un mot

- soit un mot $w \in \Sigma^*$,
- soit un ensemble $X$ de facteurs dans $w$,
- existe t'il un algorithme polynomial pour tester si $X$ partitionne $w$ ?

# X-factorisation d'un mot

**Théorème[Rivière, Barth, Cohen, Denise]**

Soit $X$ un ensemble de mots et $w$ un mot. Le problème de décider si $X$ partitionne $w$ est NP-complet.