جامعة المنارة

كلية الهندسة

قسم المعلوماتية

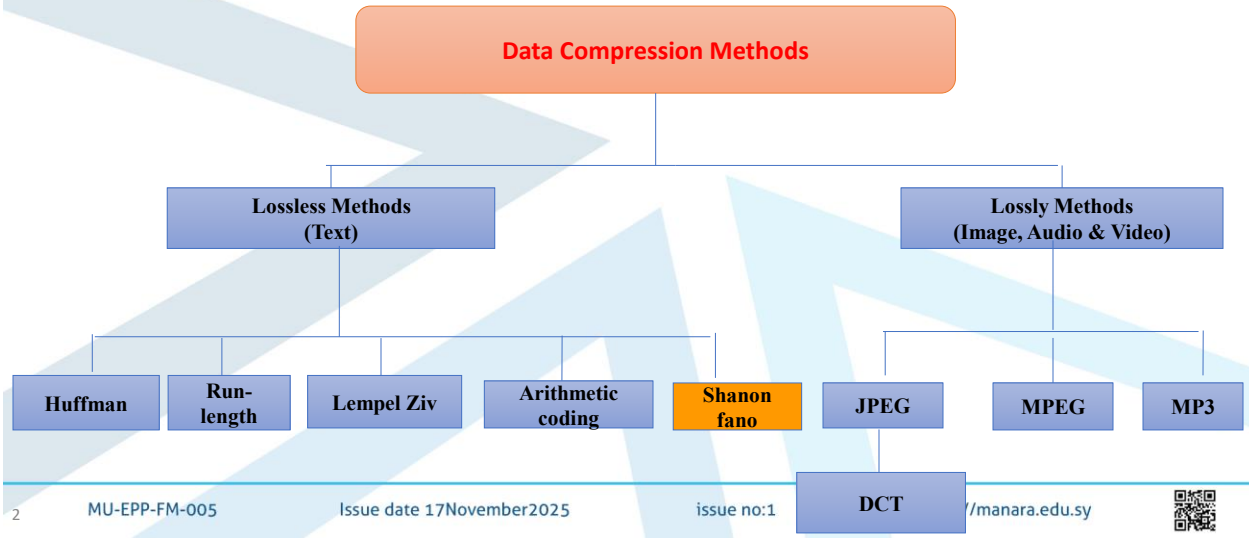
Information theory

نظرية المعلومات

مدرسة المقرر

د.بشرى علي معلا

طرائق ضغط البيانات



تعريف ترميز شانون-فانو (Shanon-Fano Definition)

- ✓ طور العالمان Claude E.Shannon وRobert M.Fano إجرائية ترميز ترتكز في عملها على تجميع مجموعة رموز مع الاحتمالات المتعلقة بها.
- ✓ يعد أول طريقة مطورة لإيجاد ترامیز ذات طول متغير.
- ✓ يعد سهل التنفيذ وذا أداء عالٍ مقارنة مع المتطلبات البسيطة لعملية البرمجة.
- ✓ لا يندرج تحت عنوان الترامیز ذات الأهمية الكبيرة كون فعالية ترميزه منخفضة مقارنة مع الترامیز الشهيرة كترميز هوفمان.
- ✓ يمكن القول أنه لا يقدم ترامیز منبع مثالية لذا فهو قليل الاستخدام ويفضل استخدام ترميز هوفمان بدلاً منه.
- ✓ إن هذا الترميز يحقق نتائج أفضل عندما ترد الرموز وفق احتمالات تكرار مرفوعة للقوة 2.

مقارنة مع ترميز هوفمان

- ✓ تعد طريقة ترميز شانون-فانو مشابهة بعض الشيء لترميز هوفمان
  - فهي تنتج الترامیز الأفضل المتناسبة مع احتمالات التكرار.
  - تحقق الترميز الأمثل عندما تكون احتمالية الرموز مرفوعة للقوة 2
- ✓ الاختلاف الأساسي بين هاتين الطريقتين هي أن شانون- فانو يبني الترميز من الأعلى للأسفل، بينما يشكل هوفمان شجرة الترميز بطريقة معاكسة.

آلية العمل

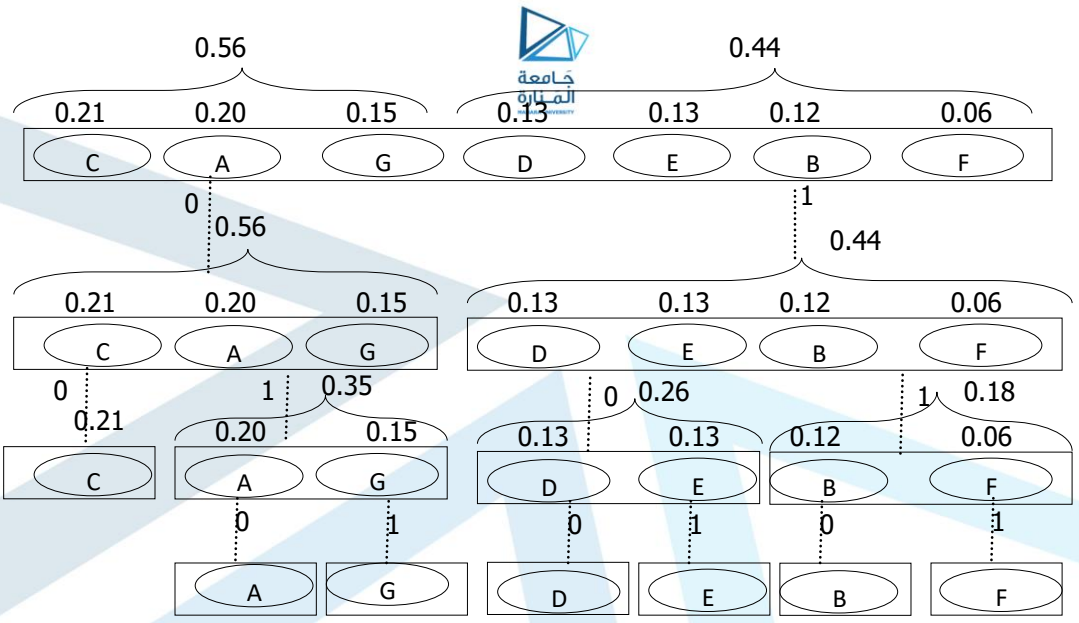
- نشكل قائمة متناسبة مع احتمالات أو تكرار الرموز.
- نرتب هذه الرموز وفقاً لتكرارها ضمن مجموعة بحيث تكون الرموز الأكبر احتمالاً في اليسار، والرموز الأقل احتمالاً في اليمين.
- نقسم هذه المجموعة لجزأين، مع عدد احتمال في الجزء اليساري أقرب ما يمكن للاحتمال في الجزء اليميني.
- نخصص للجزء اليساري الرمز **0** وللجزء اليميني الرمز**1**.
- نكرر الخطوتين الأخيرتين على كل مجموعة.
- ونستمر بذلك حتى نصل إلى نهاية الشجرة.
- مع ملاحظة أنه كلما كان التقسيم لمجموعات جزئية أفضل (الاحتمالات بين هذه المجموعات متقاربة جداً)، كلما حصلنا على الترميز الأفضل والأكثر مثالية.

مثال (١)

ليكن لدينا النص الآتي المكون من مجموعة من المحارف مع احتمالاتها والتي يريد المرسل إرسالها كما هو موضح بالجدول:

المحرف	A	B	C	D	E	F	G
الاحتمال	0.2	0.12	0.21	0.13	0.13	0.06	0.15

والمطلوب: أوجد ترميز كل محرف من محارف النص الواردة في الجدول السابق باستخدام ترميز شانون-فانو؟



ترميز شانون-فانو:فيكون ترميز كل محرف من محارف النص الواردة في الجدول السابق باستخدام

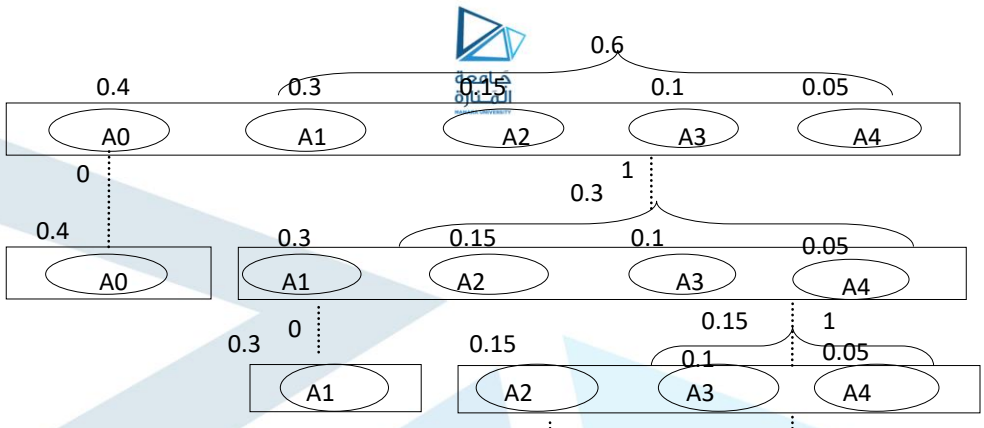
المحرف	A	B	C	D	E	F	G
الاحتمال	0.2	0.12	0.21	0.13	0.13	0.06	0.15
ترميزه	010	110	00	100	101	111	011

مثال 2

يولد مصدر معلومات A الرموز الآتية {A0,A1,A2,A3,A4} وفق الاحتمالات الآتية :

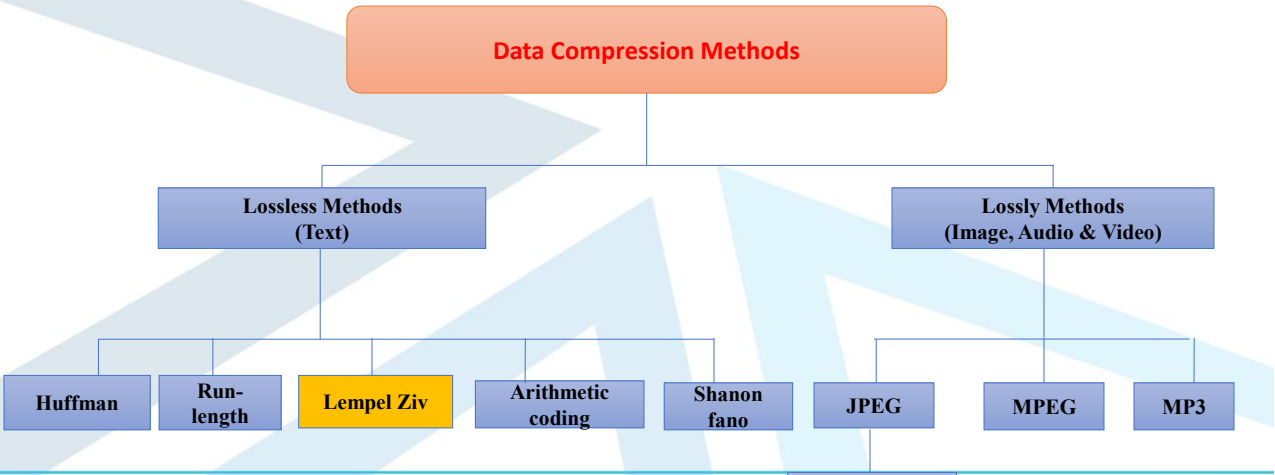
A0=0.4, A1= 0.3, A2=0.15, A3=0.1, A4= 0.05

والمطلوب:أوجد ترميز شانون-فانو لهذه الرموز.



المحرف	A0	A1	A2	A3	A4
الاحتمال	0.4	0.3	0.15	0.1	0.05
ترميزه	0	10	110	1110	1111

طرائق ضغط البيانات



Lempel-Ziv Algorithm

- تعرف هذه الخوارزمية بـ Dictionary Coding.(خوارزميات الضغط عن طريق الإستبدال النصي أو الضغط عن طريق القواميس)
- هي خوارزمية ضغط دون ضياع، وتعد أحد أصناف الترامیز الإحصائية
- تعتمد على مراقبة العلاقة بين أجزاء البيانات الواردة واستبدال جميع الكلمات المكررة بدليل يشير إليها وهو أقصر منها وذلك في جدول dictionary الذي يحوي الكلمات الواردة في النص المراد ترميزه.
- تصنف خوارزميات الضغط (LZ) إلى صنفين:

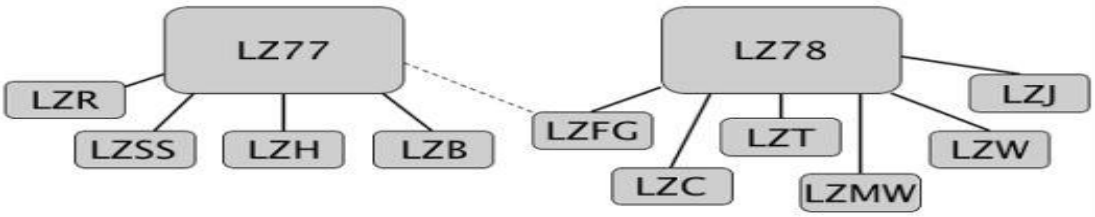
i. LZ77

ii. LZ78

تتفق الخوارزميتين من حيث المبدأ.

Lempel-Ziv Algorithm

تصنف خوارزميات الضغط (LZ) إلى صنفين:



Applications:
- zip
- gzip
- Stacker
- ...

Applications:
- GIF
- V.42
- compress
- ...

الخوارزمية LZ78

مبدأ العمل:

تعتمد هذه الخوارزمية على بناء جدول يتضمن كل الكلمات
الواردة في النص لمرة واحدة (أول مرة)،
بحيث لا يكون هناك حاجة عند الترميز لإرسال الكلمة عند ورودها في النص مرة أخرى
وإنما يرسل **دليلاً يشير إلى موقع هذه الكلمة ضمن الجدول**.

الخوارزمية LZ78

➢ عملية بناء القاموس dictionary

• في حال ورود الكلمة لأول مرة:

ترمز بكلمة ترميز وتكون مؤلفة من جزأين هما الدليل والكلمة نفسها : (index,word)

➢ **في حال كانت الكلمة في الجدول:**

هنا لا داعٍ لإرسال الكلمة ويتم إرسال الدليل فقط كالتالي **(index,)**

(0, char)	if one-character pattern is not in Dictionary.
(DictionaryPrefixIndex, lastPatternCharacter)	if multi-character pattern is not in Dictionary.
(DictionaryPrefixIndex,)	if the last input character or the last pattern is in the Dictionary.

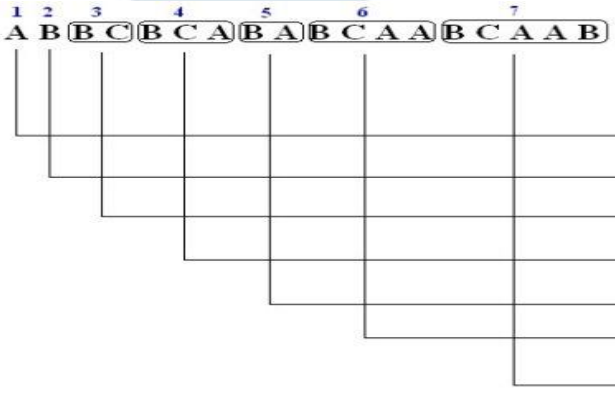
الخوارزمية LZ78

يعرف هذا الجدول بالقاموس المتكيف (adaptive dictionary) :

أي أنه لا حاجة لأن يكون فاك الترميز على **علم مسبق** بمحتويات هذا الجدول، وإنما يقوم **ببناء جدوله حالما يستقبل البيانات** الواردة وذلك بآلية مشابهة للآلية التي يعمل بها المرمز

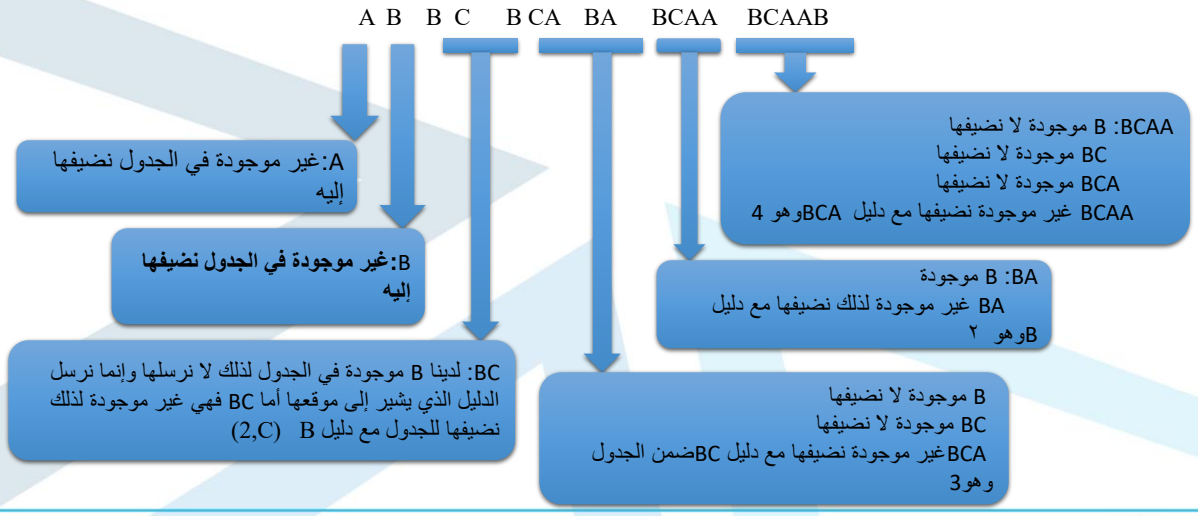
مثال (١)عن الخوارزمية LZ78

المطلوب ترميز سلسلة المحارف الآتية: **ABBCBCABABCAABCAAB**



	Dictionary	
output	index	string
(0, A)	1	A
(0, B)	2	B
(2, C)	3	BC
(3, A)	4	BCA
(2, A)	5	BA
(4, A)	6	BCAA
(6, B)	7	BCAAB

مثال : ABBCBCABABCAABCAAB



مثال (١)عن الخوارزمية LZ78

- يرمز الجزء الصحيح من كلمة الترميز بعدد بتات تمثل الترميز الثنائي لـ (للدليل –1)

الرسالة المرمزة: (0,A)(0,B)(2,C)(3,A)(2,A)(4,A)(6,B)

1 2 3 4 5 6 7

- مثلاً (4,A)حسب الرسالة المرمزة دليل كلمة الترميز هذه هو 6 و الجزء الصحيح هو 4:

Index=6 ⟵ Index-1=5 ⟵ لترميز العدد 5 إلى 3 خانات ⟵ 4 ترمز بـ 100

- و منه: الرسالة المرمزة: (0,A)(0,B)(2,C)(3,A)(2,A)(4,A)(6,B)

index 1 2 3 4 5 6 7

0A0B10C11A010A100A110B

مثال (١)عن الخوارزمية LZ78

- مقدار الضغط =عدد البتات المطلوبة لترميز هذه الرسالة بالآسكي مقسوماً على عدد البتات اللازمة للترميز باستخدام الخوارزمية LZ78

- يكون عدد البتات المطلوبة لترميز هذه الرسالة بلغة ASCIIهو:

7*18=126 bits

- يكون عدد البتات المطلوبة للترميز باستخدام الخوارزمية LZ78 :

0A 0B 10C 11A 010A 100A 110B

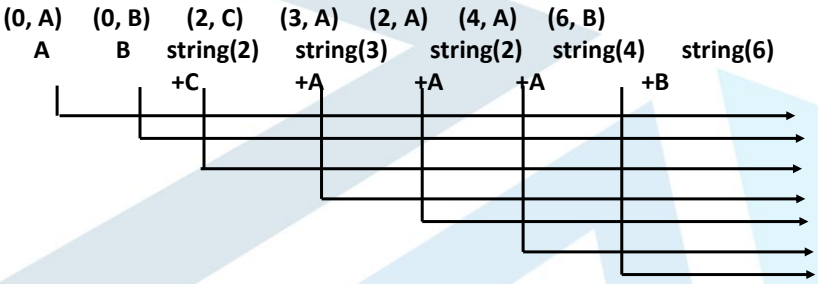
(1+7) (1+7) (2+7) (2+7) (3+7) (3+7) (3+7)

(1+7) + (1+7) + (2+7) + (2+7) + (3+7) + (3+7)+ (3+7)= 64 bits

- مقدار الضغط: 126/64=1.96

مثال (٢)عن الخوارزمية LZ78

المطلوب: فك ترميز الرسالة : (0, A) (0, B) (2, C) (3, A) (2, A) (4, A) (6, B)
1 2 3 4 5 6 7



	Dictionary	
Output	Index	input
A	1	(0,A)
B	2	(0,B)
BC	3	(2,C)
BCA	4	(3,A)
BA	5	(2,A)
BCAA	6	(4,A)
BCAAB	7	(6,B)

- الرسالة بعد فك ترميزها: **ABBCBCABABCAABCAAB**

إيجابيات وسلبيات الخوارزمية LZ

إيجابيات الخوارزمية lempl zif

- ❖ تحقق نسبة ضغط كبيرة وفعالة من أجل النصوص الكبيرة
- ❖ لا داع لأن يحتوي فاك الترميز على (dictionary) مسبقاً، لأنه يقوم ببنائه اعتماداً على البيانات الواردة بالتتالي

سلبيات الخوارزمية lempl zif

- ❖ غير فعالة في ضغط النصوص الصغيرة

جَـامعة
المَـنارة
MANARA UNIVERSITY

نهاية المحاضرة الخامسة