

07_환경 분석으로 빅데이터의 서비스를 시작하라!

#1

1. 빅데이터 서비스 환경 분석하기

가. 조직 내·외부의 환경 특징 분석 도구

조직의 경영 전략, 마케팅에 대한 조직 내·외부의 환경 특징을 분석하는 도구로서 일반적으로 3C(Customer, Company, Competitor), SWOT(Strength, Weakness, Opportunity, Threat), BCG 분석 등을 사용합니다. 이와 같은 환경 분석 도구를 활용해 서비스 모델과 관련한 내·외부의 환경 분석을 수행합니다.

#2

① 3C 분석

3C(Customer, Company, Competitor)는 경쟁사(경쟁 조직)와의 비교를 통해 서비스 모델의 내·외부 환경 특징을 분석하는 도구입니다. 고객과 경쟁사 분석을 통한 외부 환경 분석을 실시하고, 기업(조직) 분석을 통한 내부 환경 분석을 수행합니다.

② SWOT 분석

SWOT 분석은 주로 조직의 경영 전략을 도출하는 도구로 사용됩니다. 빅데이터 서비스 모델과 관련하여 내·외부 환경 분석에 사용할 수 있습니다. 외부 환경에 대한 분석을 통해 기회(Opportunity)와 위협(Threat)을 파악하고, 내부 현황에 대한 분석을 통해 강점(Strength)과 약점(Weakness)을 파악합니다. 이를 바탕으로 조직의 경영 전략을 도출해 낼 수 있습니다.

#3

③ BCG 분석

조직의 현재 상황을 파악할 때 주로 사용하는 환경 분석 도구입니다. 처음 이 도구를 소개한 보스턴 컨설팅 그룹(Boston Consulting Group)의 이름을 따서 'BCG 매트릭스(Matrix)'라고 부르기도 합니다. 시장 점유율(Market Share)과 시장 성장률(Market Growth)에 대해 네 가지로 구분하여 환경 분석을 진행합니다.

㉓ 별(Star)

- 시장의 성장률과 점유율이 모두 높은 서비스 모델입니다.
- 지속적으로 많은 투자와 관리가 필요합니다.
- 현금을 많이 벌어들이지만 재투자가 필요합니다.

㉔ 물음표(Question Mark)

- 시장의 성장률은 높지만 시장 점유율은 낮은 서비스 모델입니다.
- 투자를 확대해 서비스의 성장을 노릴 것인지, 아니면 시장을 포기할 것인지 의사결정이 필요합니다.

㉕ 캐시 카우(Cash Cow)

- 시장의 성장률은 낮지만 시장 점유율이 높은 서비스 모델입니다.
- 안정적 수익과 현금 흐름의 개선이 가능합니다.
- 추가 투자보다는 현 상황의 시장 점유율을 유지해야 합니다.

㉖ 개(Dog)

- 시장의 성장률과 시장 점유율이 모두 낮은 서비스 모델입니다.
- 쇠퇴기에 들어간 서비스 모델이며 재투자를 지양해야 합니다.

#4

2. 빅데이터 데이터의 환경 분석하기

가. 빅데이터 관련 소프트웨어의 공학 방법론

빅데이터 서비스 모델을 개발하기 위한 소프트웨어 공학적 방법론은 대표적으로 네 가지가 있습니다.

- ① 빅데이터 개발 방법론: 일반적으로 빅데이터 시스템에 대한 개발 구현과 빅데이터 분석 모델 방법론으로 구성됩니다.
- ② 데이터 표본 추출 방법론: 데이터의 표본조사와 추출에 대한 방법을 정의합니다.
- ③ 데이터 모델링 방법론: 데이터의 이해와 준비, 탐색과 실증 분석을 통한 데이터의 모델링을 수행합니다.
- ④ 데이터 품질 관리 방법론: 서비스 모델의 데이터에 대한 품질 관리와 성숙도를 관리합니다.

#5

나. 데이터 프로파일링(Data Profiling)의 개념

데이터의 품질 진단을 위해 사용하는 기법 중의 하나입니다. 데이터베이스(DB: Data Base) 내의 데이터를 읽어 테이블(Table)이나 칼럼(Column)에 대한 데이터의 현황 정보를 통계적으로 분석하는 것입니다. 일반적으로 프로파일링은 테이블 단위와 칼럼 단위로 구분하게 됩니다.

- ① 메타데이터의 수집: 프로파일링의 기준이 되는 메타데이터를 수집합니다.
- ② 대상의 선정: 테이블 단위 또는 칼럼 단위로 대상을 선정합니다.
- ③ 프로파일링의 수행: 유일성 측면의 중복 분석, 일관성 측면의 관계 분석, 데이터 카테고리화 유형 분석, 데이터의 빈도와 평균 및 편차 등을 분석합니다.
- ④ 결과 검토 및 확정: 프로파일링 결과에 대한 검토와 프로파일링의 결과를 확정합니다.

#6

다. 개인정보 보호를 위한 비식별 처리

① 데이터의 라이프 사이클

데이터의 라이프 사이클 과정상의 비식별 처리는 개인 프라이버시에 영향을 주지 않으면서 개인 데이터를 사용하기 위한 목적으로 수행됩니다.

- 데이터의 생성: 데이터를 생성할 때 개인 식별 정보를 삭제합니다.
- 데이터의 수집: 데이터가 수집되지만 개인 식별 정보가 필요하지 않은 경우와 데이터를 수집하는 경로에서 비식별 처리를 합니다.
- 데이터의 저장 및 처리: 개인 식별 정보에 대한 관리를 회피하기 위한 수단으로 데이터를 변환한 뒤에 또는 데이터를 저장하기 전에 비식별 처리를 합니다.
- 데이터의 활용: 분석 용도로 개인정보를 이용한 뒤에 활용 전의 데이터를 비식별 처리합니다.

#7

② 데이터의 비식별 처리

데이터 활용 모델을 통한 비식별 처리는 재식별 가능성을 줄이기 위해 계층적 접근 제어 등을 활용합니다.

㉓ 공개적 활용 모델

- 모든 사람 또는 조직은 데이터에 접근이 가능합니다.
- 무제한 사용과 재사용이 가능합니다.
- 재식별 위험이 매우 크며, 모든 재식별 공격의 위험성이 있습니다.

㉔ 비공개적 활용 모델

- 특정 사람 또는 조직에 한하여 데이터에 접근이 가능합니다.
- 데이터의 재활용이 금지되어 있습니다.
- 제한을 엄격하게 두는 경우 재식별의 위험은 보통 수준입니다.
- ‘조직 내부의 의도적 공격’과 ‘데이터 유출’의 재식별 공격의 위험성이 있습니다.

㉕ 반공개적 활용 모델

- 인가된 사람 또는 조직은 데이터에 접근이 가능합니다.
- 인가된 사람 또는 조직에 한하여 사용하며, 재사용이 가능합니다.
- 재식별 위험이 매우 큼니다.
- ‘조직 내부의 의도적 공격’과 ‘데이터 유출’의 재식별 공격의 위험성이 있습니다.

#8

질문자: 빅데이터의 정보 보안 문제를 확인하기 위해 어떤 사항을 고려해야 하나요?

전문가: 선정된 데이터에 대한 양(Volume), 속도(Velocity), 다양성(Variety)을 파악하고 이를 기존의 보안 솔루션이 지원할 수 있는지 확인합니다. 동일한 암호화 기술을 적용할 수 있다면 데이터의 대용량 및 다양성으로 인한 성능, 비용 문제를 해결해야 합니다. 또 빅데이터 규제 사항 준수와 분석 효과를 최대한 보장할 수 있도록 정보 보안의 방안을 마련합니다. 그리고 데이터 품질 및 정확성을 유지한 상태에서 정보 보안 정책을 수립해야 합니다.

#9

3. 빅데이터 기술 환경 분석

가. 빅데이터 데이터 전·후 처리 기법

빅데이터의 데이터 전·후 처리 방식에는 데이터 정제, 데이터 축소, 데이터 통

합, 데이터 변환 등을 수행합니다. 데이터 전·후 처리를 하지 않으면 빅데이터 분석에서 의미 있는 결과가 나오기 어렵습니다.

#10

① 데이터 정제

데이터 이상치(Outlayer), 데이터 결측값(Missing Value: 빠진 데이터) 등에 대한 보정 작업을 통해 데이터 신뢰도를 향상합니다. 사람의 실수, 데이터 표현의 불일치, 의도와 다른 목적에 사용되었는지에 관련하여 오류사항을 포착합니다.

② 데이터 축소

대량의 데이터의 크기를 축소하여 빅데이터의 분석을 진행합니다. 크기를 축소하기 전과 후는 동일한 분석 결과가 나올 수 있어야 합니다. 데이터 표본을 추출하거나 통계 기법을 적용할 수 있습니다. 분석의 목적에 맞는 데이터 및 속성을 추출합니다.

#11

③ 데이터 변환

데이터 단위의 변환, 데이터 편차 등에 대한 표준 편차를 적용하고 변환 또는 변형 작업을 진행합니다. 데이터 단위의 변환, 데이터 키 또는 값을 매핑하거나 변환합니다.

④ 데이터 통합

다수의 데이터를 연계하고 통합하는 작업을 수행합니다. 동일한 뷰(View)에서 볼 수 있도록 통합하는 작업을 수행합니다. 데이터 및 스키마를 통합하고 개체를 식별합니다. 또 데이터 값의 충돌을 해결해야 합니다.

#12

나. 빅데이터 분석 결과 시각화를 위한 그래픽 단순화

그래픽 단순화는 빅데이터 분석 결과를 시각화하기 위하여 사용합니다. 그래픽 단순화는 범주와 차원의 수에 따라 선택된 그래프에 대해 최소한의 표현으로 최대한의 것을 전달하고 표현하는 것을 의미합니다. 그래픽 단순화를 위해 시각화를 위한 디자인의 기본 원리를 파악합니다.

① 타이포그래피

- 2가지 이하의 서체에 크기나 스타일에 변화를 주어야 합니다.
- 정보의 차별화나 강조 등을 위해 선택적으로 사용합니다.

② 색상

- 두 가지의 색을 사용할 때 보색을 이용합니다. 채도와 명도는 동일하게 처리합니다.
- 색채를 사용할 때 인간의 지각 및 인식이 작용됩니다. 예를 들면 붉은색은 금지를 의미하고, 초록색은 안전을 의미하는 것으로 표시할 수 있습니다.

#13

③ 그리드

- 디자인 내부에 여러 요소를 표현할 때 그리드에 대한 배열을 계획합니다.
- 그리드 내부에 대해 블록으로 나누어 효율적으로 요소를 배치합니다.
- 사람의 눈은 왼쪽 상단에서 오른쪽 하단으로 시각이 이동한다는 점을 활용합니다.

④ 아이소타입(ISOTYPE)

- 아이소타입(ISOTYE)은 국제 그림글자 교육기구(International System of Typographic Picture Education)를 의미합니다.
- 국제적인 그림 언어 체계로 각종 지식에 대한 조직적 시각화가 가능합니다.
- 문자와 숫자를 사용하는 대신에 상징적인 도형이나 정해진 기호를 조합합니다.

⑤ 상호 작용

- 웹 등의 디지털 환경에서 사용자나 고객과 인터랙션하는 것을 의미합니다.
- 웹 사이트 링크를 통해 사용자가 콘텐츠를 선택하고 상세 자료로 이동할 수 있습니다.